MARK SOLMS



EL MANANTIAL OCULTO

Un viaje a la fuente de la **conciencia**

Capitán Swing

[image]

Índice de figuras

Aunque se ha hecho todo lo posible para contactar con los titulares de los de

EL MANANTIAL OCULTO

Un viaje a la fuente de la **conciencia**

En memoria de Jaak Panksepp
(1943-2017)

Resolvió el viejo acertijo y
fue un verdadero sabio.

Introducción

Cuando era pequeño, se me ocurrió una pregunta un tanto especial: ¿cómo imaginamos el mundo tal como existía antes de que evolucionara la conciencia? Porque ese mundo existía, por supuesto, pero ¿cómo imaginarlo tal como era antes de que fuera posible imaginar algo?

Para que se hagan una idea de lo que quiero decir, intenten imaginar un mundo en el que no puede salir el sol. La Tierra siempre ha girado alrededor del Sol, pero el sol solo aparece en el horizonte desde el punto de vista de un observador. Es un acontecimiento inherentemente perspectivo. El amanecer estará atrapado para siempre en la experiencia.

Esta obligación de adoptar una perspectiva es lo que hace que nos cueste tanto comprender la conciencia. Para ello tenemos que eludir la subjetividad: mirarla desde fuera, ver las cosas como son en verdad en lugar de como se nos aparecen. Pero ¿cómo lograrlo? ¿Cómo escapar de nosotros mismos?

De joven visualizaba ingenuamente mi conciencia como una burbuja que me rodeaba: contenía los sonidos, las imágenes en movimiento y otros fenómenos de la experiencia. Fuera de la burbuja, suponía que había una negrura infinita, una negrura que imaginaba como una sinfonía de —entre otras cosas— cantidades puras, fuerzas y energías que interactuaban entre sí, etc.: la realidad auténtica de «ahí fuera» que mi conciencia representa en las formas cualitativas en las que debe hacerlo.

La imposibilidad de imaginar algo así —la imposibilidad de representar la realidad sin representaciones— ilustra la magnitud de la tarea que me impongo con este libro. Tras todos estos años, levanto otra vez el velo de la conciencia para averiguar qué vislumbro sobre su mecanismo real.

En consecuencia, el libro que tienen entre las manos es inevitablemente perspectivista. De hecho, lo es incluso más de lo que requiere la paradoja que acabo de describir. Para ayudarles a ver las cosas desde mi punto de vista, he decidido contar parte de mi propia historia. Con frecuencia, mis ideas científicas sobre la conciencia han avanzado a partir de acontecimientos de mi vida personal y mi trabajo clínico, y aunque creo que mis conclusiones se sostienen por sí solas, es mucho más fácil entenderlas si se sabe cómo he llegado a ellas.

Algunos de mis descubrimientos —por ejemplo, los mecanismos cerebrales de los sueños— se han producido en gran medida por casualidad. Algunas de mis decisiones profesionales —por ejemplo, tomar un desvío de mi carrera de neurocientífico y formarme como psicoanalista— han dado mejores frutos de los que me hubiera atrevido a esperar. Veremos lo que pasó en ambos casos.

Sin embargo, no puedo hablar del éxito de mi misión para comprender la conciencia sin señalar lo afortunado que he sido de contar con los colaboradores más brillantes. En concreto, tuve la inmensa suerte de poder trabajar con el difunto Jaak Panksepp, un neurocientífico que entendió mejor que ningún otro el origen y el poder de los sentimientos. Sus ideas conforman casi todas mis creencias actuales sobre el cerebro.

En tiempos más recientes he podido trabajar con Karl Friston, quien, entre sus muchas excelentes virtudes, puede presumir de ser el neurocientífico vivo más influyente del mundo. Fue Friston quien estableció los primeros cimientos de la teoría que voy a exponer. Se lo conoce sobre todo por reducir las funciones cerebrales (de todo tipo) a una necesidad física básica consistente en minimizar algo llamado «energía libre». Ese concepto se explica en el capítulo 7, pero, por ahora, puedo adelantar que la teoría que Friston y yo hemos desarrollado se une a ese proyecto; tanto que podríamos llamarla la «teoría de la energía libre de la conciencia». Porque de eso se trata.

La explicación definitiva de la sintiencia es un enigma tan complicado de resolver que hoy día es conocido reverencialmente como «el problema difícil». A veces, cuando se ha resuelto un enigma, tanto la pregunta como la respuesta pierden su interés. Dejo a los lectores decidir si las ideas que expondré aquí arrojan nueva luz sobre el problema difícil. En cualquier caso, confío en que nos ayudarán a vernos a nosotros mismos bajo una nueva luz, y solo por eso deberían ya conservar su interés hasta el momento en que sean sustituidas. A fin de cuentas, en un sentido profundo, todos somos nuestra conciencia. Parece, pues, razonable esperar que una teoría de la conciencia explique los fundamentos de la razón por la cual nos sentimos como nos sentimos. Debería explicar por qué somos como somos. Tal vez incluso debería aclarar qué podemos hacer al respecto.

Reconozco que este último tema queda fuera del alcance previsto para este libro, pero no del alcance de la teoría. Mi relato de la conciencia aúna en una sola historia la física elemental de la vida, los avances más recientes tanto de la neurociencia computacional como de la afectiva y las sutilezas de la experiencia subjetiva que

tradicionalmente ha explorado el psicoanálisis. Dicho de otra forma, la luz que arroja esta teoría debería poder iluminarnos a todos.

Ha sido el trabajo de mi vida. Décadas después, sigo preguntándome cómo se vería el mundo antes de que hubiera nadie para verlo. Ahora, más instruido, imagino el origen de la vida en una de esas fuentes hidrotermales. Seguramente, los organismos unicelulares que empezaron a existir allí no serían conscientes, pero sus perspectivas de supervivencia debieron de verse afectadas por lo que los rodeaba. Es fácil imaginar aquellos organismos sencillos respondiendo a la «bondad» biológica de la energía del sol. A partir de ahí, basta con dar un pequeño paso para imaginar seres más complejos buscando de forma activa suministros de energía y desarrollando al final una capacidad de sopesar las posibilidades de éxito de distintas alternativas.

La conciencia, a mi entender, surgió de la experiencia de aquellos organismos. Imaginen el calor del día y el frío de la noche desde la perspectiva de aquellos primeros seres vivos. Los valores fisiológicos que registraron sus experiencias diurnas fueron los precursores del primer amanecer.

Muchos filósofos y científicos siguen creyendo que la sintiencia no tiene utilidad física alguna. Mi objetivo con este libro es persuadir a los lectores de la verosimilitud de otra interpretación, y para ello tengo que convencerles de que los sentimientos forman parte de la naturaleza, que no son en esencia distintos de otros fenómenos naturales y que tienen un papel dentro de la matriz causal de las cosas. Demostraré que la conciencia tiene que ver con los sentimientos, y que los sentimientos tienen que ver a su vez con lo bien o lo mal que nos va en la vida. La conciencia existe para ayudarnos a que nos vaya mejor.

Dicen que el problema difícil de la conciencia es el enigma más grande sin resolver de la neurociencia contemporánea, cuando no de toda la ciencia. La solución propuesta en este libro se aparta radicalmente de los abordajes convencionales. Dado que la inteligencia reside en la corteza cerebral, casi todo el mundo piensa que también es allí donde reside la conciencia, pero yo no estoy de acuerdo; la conciencia es mucho más primitiva: surge de una parte del encéfalo que los humanos comparten con los peces. Ese es el «manantial oculto» del título.

La conciencia no debería confundirse con la inteligencia. Es perfectamente posible sentir dolor sin la menor reflexión sobre qué es el dolor. De la misma manera, las ganas de comer —la sensación de hambre— no implican comprensión intelectual de las exigencias de la vida. La conciencia en su forma elemental, es decir, el sentimiento puro, es una función que sorprende por su sencillez.

Este enfoque lo han adoptado otros tres destacados neurocientíficos: Jaak Panksepp, Antonio Damasio y Björn Merker. Panksepp abrió el camino. Él (como Merker) investigaba con animales; Damasio (como yo), no. A muchos lectores les horrorizarán los hallazgos de la investigación con animales que expongo en el libro, precisamente porque demuestran que otros animales sienten igual que nosotros. Todos los mamíferos pueden sentir dolor, miedo, pánico, tristeza… Por irónico que parezca, fue la investigación de Panksepp la que acabó con cualquier duda razonable al respecto. Nuestro único consuelo es que sus descubrimientos frenaron ese tipo de investigaciones.

Mi atracción por Panksepp, Damasio y Merker se debe a su creencia, que comparto, de que a la neurociencia actual le falta un enfoque claro de la naturaleza intrínseca de la experiencia vivida. Se podría decir que lo que nos une es lo que hemos construido, a veces sin saberlo, sobre los cimientos abandonados que dejó Freud para una ciencia de la mente que priorice los sentimientos sobre la cognición. (La cognición es en su mayor parte inconsciente). Este es el segundo desvío radical de este libro: nos devuelve al «Proyecto» de Freud de 1895… e intenta acabar el trabajo, aunque sin pasar por alto sus muchos errores. Para empezar, Freud creía, como todos los demás, que la conciencia era una función cortical.

El tercer gran desvío que toma este libro es llegar a la conclusión de que la conciencia es modelable, artificialmente producible. Esta conclusión, con sus profundas implicaciones metafísicas, surge de mi trabajo con Karl Friston. A diferencia de Panksepp, Damasio y Merker, Friston es un neurocientífico computacional. Por consiguiente, cree que la conciencia se puede reducir en última instancia a las leyes de la física (una creencia que, sorprendentemente, compartía Freud). Sin embargo, antes de que empezáramos a colaborar, el propio Friston equiparaba en gran medida las funciones mentales con las corticales. Este libro profundiza en su marco estadístico-mecánico, adentrándose en los recovecos más primitivos del tronco encefálico…

Estos tres desvíos hacen menos difícil el problema difícil. Y el presente libro explica cómo.

M

ARK

S

OLMS

Chailey, East Sussex

Marzo de 2020

01

La materia de los sueños

Nací en la Costa de los Esqueletos, en la antigua colonia alemana de Namibia, donde mi padre administraba una pequeña empresa sudafricana, la Consolidated Diamond Mines. Su sociedad matriz, De Beers, había creado prácticamente un país dentro del país, el llamado Sperrgebiet (la «zona prohibida»). Sus minas de aluvión se desparramaban desde las dunas del desierto de Namib hasta el fondo del océano Atlántico, varios kilómetros mar adentro.

Ese fue el peculiar paisaje que moldeó mi imaginación. De pequeño, solía jugar con mi hermano mayor Lee a extraer diamantes, recreando en nuestro jardín, con excavadoras de juguete, las impresionantes obras de ingeniería que veíamos cuando nuestro padre nos llevaba a visitar las minas a cielo abierto en medio del desierto. (Ni que decir tiene que éramos demasiado pequeños para conocer los aspectos menos admirables de su industria).

Un día de 1965, cuando yo tenía cuatro años, mis padres habían salido a navegar desde el club náutico Cormorant, como muchas otras veces, y a mí me dejaron jugando en el club con Lee, que tenía seis años. Ya se había disipado la bruma matutina y salí del fresco interior del edificio de tres pisos que albergaba el club para acercarme a la orilla. Allí, caminando por el agua en medio de aquel calor, contemplé cómo se dispersaban alrededor de mis pies unos pececillos minúsculos y brillantes, mientras Lee y unos amigos trepaban al tejado del edificio por la parte de atrás.

Lo que recuerdo a partir de ahí son tres fotos fijas. La primera, el sonido de algo que se partió como una sandía. La segunda, la imagen de Lee tumbado en el suelo, gimiendo por lo mucho que le dolía una pierna. La tercera, mis tíos diciéndome que cuidarían de mi hermana y de mí mientras nuestros padres llevaban a Lee al hospital. Los lamentos por el dolor en la pierna deben de ser una fabulación, porque la historia clínica dice que mi hermano perdió el conocimiento al golpearse contra el suelo de hormigón.

Puesto que Lee necesitaba una atención especializada que nuestro hospital local no podía prestar, lo trasladaron en helicóptero al hospital Groote Schuur de Ciudad del Cabo, a ochocientos kilómetros de distancia. En aquella época, el Departamento de Neurocirugía se

hallaba en un impresionante edificio construido según el estilo arquitectónico neerlandés del Cabo, el mismo en el que ahora trabajo como neuropsicólogo. Lee había sufrido una fractura de cráneo con hemorragia intracraneal. Cuando este tipo de hematomas se extiende, suponen un riesgo para la vida del paciente y requieren intervención quirúrgica. Mi hermano tuvo suerte; los hematomas le desaparecieron en los días siguientes y pronto le dieron el alta.

Al margen del casco que tuvo que llevar tras el accidente para proteger el cráneo fracturado, Lee parecía el de siempre. Sin embargo, como persona había cambiado por completo. Hay una palabra alemana que describe muy bien lo que sentí, Unheimlichkeit, para la que no encuentro equivalente, pero que se podría traducir como «inquietante extrañeza».

El cambio más evidente fue la pérdida de sus hitos del desarrollo. Durante un tiempo perdió incluso el control de esfínteres. A mí lo que más me inquietaba era que parecía que ya no pensaba igual que antes. Era como si Lee estuviera y, al mismo tiempo, no estuviera. Había olvidado muchos de nuestros juegos. El de extraer diamantes pasó a ser simplemente cavar agujeros. Mi hermano ya no conectaba con sus aspectos imaginativos y simbólicos. Ya no era Lee.

Aquel año suspendió en la escuela, por primera vez. Lo que más recuerdo de aquellos primeros días tras el accidente son mis intentos de conciliar la dicotomía de que el hermano que había regresado parecía el mismo pero no lo era. Me preguntaba adónde había ido a parar su versión anterior.

En los años siguientes caí en una depresión. Recuerdo que por las mañanas no podía reunir la energía necesaria para calzarme e ir a la escuela. Habían pasado unos tres años desde el accidente. No encontraba la energía para hacer nada porque no le veía el sentido. Si nuestro ser dependía del funcionamiento de nuestro cerebro, ¿qué ocurriría conmigo cuando mi cerebro muriera? ¿Y con el resto de mi cuerpo? Si la mente de Lee podía reducirse a un órgano, seguro que la mía también. Eso significaba que yo —mi ser sintiente— solo existiría durante un periodo de tiempo relativamente breve. Y luego desaparecería.

He pasado toda mi carrera científica reflexionando sobre este problema. Quería entender qué le había pasado a mi hermano y qué nos pasaría a todos con el tiempo. Necesitaba entender en qué consistía, en términos biológicos, nuestra existencia como sujetos que experimentan. En pocas palabras, entender la conciencia. Por eso me

hice neurocientífico.

Incluso visto en retrospectiva, creo que no podía haber tomado un camino más directo para dar con las respuestas que buscaba.

Se podría afirmar que la naturaleza de la conciencia es el tema más difícil de la ciencia. Un tema que importa porque cada cual es su propia conciencia, pero que es controvertido por culpa de dos enigmas que han frustrado a los pensadores durante siglos. El primero es el modo en que la mente se relaciona con el cuerpo o, para los investigadores con una perspectiva más materialista —o sea, para la mayoría de los neurocientíficos—, la manera en que el cerebro da lugar a la mente, es decir, al llamado «problema mente-cuerpo». ¿Cómo produce el cerebro físico nuestra experiencia fenoménica? O, con el mismo nivel de confusión, ¿cómo la materia no física llamada «conciencia» controla el cuerpo físico?

Los filósofos han asignado este problema a lo que ellos llaman «metafísica», que es una forma de decir que no creen que se pueda resolver científicamente. ¿Por qué no? Porque la ciencia depende de métodos empíricos, y «empírico» implica «derivado de indicios sensoriales». La mente no es accesible a la observación sensorial: no se puede ver ni tocar; es invisible e intangible; es un sujeto, no un objeto.

El segundo enigma consiste en averiguar qué podemos saber de la mente desde el exterior o, ya puestos, cómo podemos saber siquiera si está presente. Es el denominado «problema de las demás mentes». Simplificando: si la mente es subjetiva, solo se puede observar la propia. Entonces, ¿cómo podemos saber si otras personas (o animales o máquinas) la tienen siquiera? O, aún más, ¿cómo vamos a deducir qué leyes objetivas rigen el funcionamiento general de nuestra mente?

En el siglo pasado, estas preguntas dieron lugar a tres grandes respuestas científicas. La ciencia se basa en los experimentos. A nuestro favor tenemos que el método experimental no aspira a alcanzar verdades definitivas, sino a lo que podríamos definir como mejores suposiciones. A partir de las observaciones, proponemos conjeturas respecto a lo que podría explicar los fenómenos observados. Dicho de otro modo, formulamos hipótesis. Luego, a partir de nuestras hipótesis, generamos predicciones: «Si la hipótesis X es correcta, debería producirse Y cuando yo haga Z» (sin descartar una posibilidad razonable de que Y no se produzca según otras hipótesis). Eso es el experimento. Si no se produce Y, entonces se deduce que X es falso y

se revisa conforme a las nuevas observaciones. Después se reinicia el proceso experimental, hasta que da lugar a predicciones falsables que se confirman. Llegados a ese punto, consideramos la hipótesis provisionalmente cierta, hasta que nuevas observaciones la contradigan. Así, en la ciencia no esperamos llegar a la certeza; solo aspiramos a menos incertidumbre.[1]

En la primera mitad del siglo XX, una escuela de psicología llamada «conductismo» empezó a aplicar sistemáticamente a la mente el método experimental. Su punto de partida consistía en descartar todo lo que no fueran hechos observables empíricamente. Los conductistas desterraron todo discurso «mentalista» sobre creencias e ideas, sobre sentimientos y deseos, y restringieron su campo de estudio a las respuestas visibles y tangibles del sujeto a estímulos objetivos. Su desinterés por cualquier descripción subjetiva de lo que ocurría en el interior era casi fanático. Trataban la mente como una «caja negra» de la que solo se podía conocer lo que entraba y lo que salía.

¿Por qué adoptaron una actitud tan extrema? En parte, está claro, para esquivar el problema de las demás mentes. Negándose de entrada a todo debate sobre la mente, conseguían que sus teorías no se vieran afectadas por las dudas filosóficas endémicas de la psicología. Sí, en efecto, excluyeron la psique de la psicología.

Podría parecer un precio muy alto, pero el conductismo fue desde el principio una doctrina revolucionaria. Los conductistas no perseguían la pureza epistemológica per se: también intentaban acabar con el poder que la psicología acaparaba en aquel momento. El psicoanálisis freudiano había dominado la ciencia de la mente desde el cambio de siglo. Examinando de cerca los aspectos más curiosos de los testimonios introspectivos, Sigmund Freud pretendía desarrollar un modelo de la mente considerada, por así decirlo, desde dentro hacia fuera. Las ideas resultantes marcaron la agenda del tratamiento y la investigación durante medio siglo, y dieron lugar a la aparición de instituciones, expertos acreditados y un selecto grupo de destacados defensores intelectuales. Sin embargo, a juicio de los conductistas, todas las teorías de Freud eran meros castillos en el aire, construidos sobre los vaporosos cimientos de la subjetividad. Freud se había zambullido de cabeza en el problema de las demás mentes y había arrastrado consigo al resto de la psicología. Y correspondía a los conductistas volver a ponerla en su sitio.

Pese a la austeridad de su programa, lo cierto es que los conductistas pudieron inferir relaciones causales entre ciertos tipos de estímulos mentales y de respuestas. Y no solo eso: también lograron manipular

los inputs para provocar cambios predecibles en los outputs. Descubrieron así algunas de las leyes fundamentales del aprendizaje. Por ejemplo, cuando el desencadenante de una conducta involuntaria se asocia reiteradamente a un estímulo artificial, dicho estímulo acabará desencadenando la misma respuesta involuntaria que el estímulo innato. Así, si la visión de comida se empareja repetidamente (en animales que salivan de forma natural cuando ven comida, como los perros) con el sonido de una campana, entonces ese sonido por sí solo desencadenará la salivación; es el llamado «condicionamiento clásico». De la misma forma, si una conducta voluntaria se acompaña siempre de recompensas, esa conducta se repetirá, mientras que si se acompaña de castigos, disminuirá. Por lo tanto, si se responde con un abrazo al perro que se abalanza sobre las visitas, el animal lo hará cada vez más; si se le da un cachete, lo hará menos. Es el llamado «condicionamiento operante», también conocido como «ley del efecto».

Estos descubrimientos tuvieron especial importancia porque demostraron que la mente está sujeta a leyes naturales, como todo lo demás. Pero la mente es mucho más que el aprendizaje, y el propio aprendizaje se ve influido por más factores que los estímulos externos. Imaginemos que ahora mismo pensamos: «en cuanto acabe de leer esta página, me prepararé un té». Este tipo de pensamiento influye en nuestro comportamiento todo el tiempo. Sin embargo, los conductistas no consideraban estos informes introspectivos datos científicos aceptables, porque los pensamientos no se pueden observar desde fuera. En consecuencia, no podían saber qué es lo que nos empuja a prepararnos el té.

El gran neurólogo Jean-Martin Charcot dijo un día: «La teoría es buena, pero no impide que las cosas existan».[2] Como es evidente que los actos mentales internos existen y ejercen una influencia causal en el comportamiento, en la segunda mitad del siglo XX el enfoque conductista se vio gradualmente eclipsado por otro: la psicología «cognitiva», que, por decirlo de alguna manera, sí que podía dar cabida a los procesos mentales internos.

La revolución cognitiva se vio impulsada por la llegada de los ordenadores. Los conductistas veían el funcionamiento interno de la mente como una «caja negra» inescrutable y preferían centrarse en sus inputs y outputs. Pero los ordenadores no son insondables, y no habríamos podido inventarlos sin comprender del todo su funcionamiento interno. Por ello, al plantear la mente como si fuera un ordenador, los psicólogos se atrevieron a formular modelos del procesamiento de la información que se producía en su interior,

modelos que se pusieron luego a prueba mediante simulaciones artificiales de procesos mentales combinadas con experimentos conductistas.

¿Qué es el procesamiento de la información? Lo desarrollaré a fondo más adelante, pero lo que nos interesa más ahora es que puede realizarse con todo tipo de equipos físicos. Esto arroja nueva luz sobre la naturaleza física de la mente, porque sugiere que la mente (interpretada como procesamiento de la información) es, más que una estructura, una función. Visto así, las funciones software de la mente las realizan las estructuras hardware del cerebro, pero pueden realizarlas igual de bien otros sustratos, como los ordenadores. Así, unos y otros, cerebros y ordenadores, efectúan funciones de memoria (codifican y almacenan información) y funciones perceptivas (clasifican patrones de información entrante comparándolos con la información almacenada), además de funciones ejecutivas (ejecutan decisiones sobre qué hacer en respuesta a dicha información).

Ahí radica la fuerza de lo que se acabó llamando el «enfoque funcionalista», pero también su debilidad. Si los ordenadores —que en teoría no son seres sintientes— pueden efectuar las mismas funciones, ¿de verdad podemos reducir la mente a un mero procesamiento de la información? Hasta nuestros teléfonos móviles tienen memoria y funciones perceptivas y ejecutivas.

La tercera gran respuesta científica a la metafísica mente-cuerpo se desarrolló en paralelo a la psicología cognitiva, pero para el cambio de siglo ya había crecido hasta eclipsarla. Me refiero a un enfoque al que se le ha dado la denominación amplia de «neurociencia cognitiva». Se centra en el hardware de la mente y surgió con el desarrollo de una plétora de técnicas fisiológicas que hicieron posible observar y medir directamente la dinámica del cerebro vivo.

En la época conductista, los neurofisiólogos solo disponían de una técnica de este tipo: el electroencefalograma (EEG), que les permitía registrar la actividad eléctrica del encéfalo desde la superficie externa del cuero cabelludo. En la actualidad disponemos de muchas más herramientas, como la resonancia magnética funcional (RMF), para medir los índices de actividad hemodinámica en las distintas partes del encéfalo mientras este realiza tareas mentales concretas, o la tomografía por emisión de positrones (PET), con la que podemos medir la actividad metabólica diferencial de sistemas neurotransmisores concretos. Esto nos permite identificar con precisión qué procesos cerebrales generan nuestros distintos estados mentales. Asimismo, mediante la tractografía con tensor de difusión

podemos visualizar la conectividad funcional-anatómica detallada entre esas distintas regiones del encéfalo. Y gracias a la optogenética podemos ver y activar los circuitos de neuronas que componen las huellas mnémicas concretas a medida que se iluminan durante las tareas cognitivas.

Todas estas técnicas hacen plenamente visible el funcionamiento interno del órgano de la mente, lo que hace realidad los más atrevidos sueños empiristas de los conductistas sin limitar el ámbito de la psicología a los estímulos y las respuestas.

En los años ochenta del siglo pasado, cuando entré en el campo de la neuropsicología, esta disciplina se encontraba en un estado que explica por qué los conductistas pasaron sin gran esfuerzo de la teoría del aprendizaje a la neurociencia cognitiva. En efecto, la neuropsicología de aquella época bien podría haberse llamado «neuroconductismo». Cuanto más me enseñaban sobre funciones como la memoria a corto plazo —que decían que constituía un «búfer de datos» para mantener amarrados los recuerdos en la conciencia—, más me daba cuenta de que lo que explicaban mis profesores no era lo que yo había ido a estudiar. Nos estaban enseñando las herramientas funcionales que utiliza la mente, en lugar de la mente en sí. Me sentía muy consternado.

El neurólogo Oliver Sacks, en su libro Con una sola pierna (1984; 1994 en castellano), describió con acierto la situación en la que me vi metido:

La neuropsicología pretende ser, como la neurología clásica, completamente objetiva, y de ahí deriva su fuerza y sus avances. Pero un ser vivo, y sobre todo un ser humano, es primero y ante todo activo: un sujeto, no un objeto. Y lo que se está excluyendo es precisamente al sujeto, el «yo» vivo. La neuropsicología es admirable, pero excluye la psique, excluye al «yo» vivo, activo, que experimenta. [3]

Esa frase —«La neuropsicología es admirable, pero excluye la psique»— captaba perfectamente mi decepción. Y me llevó a entablar una correspondencia con Oliver Sacks que ya no se interrumpió hasta que falleció, en 2015. Lo que más me atrajo de Sacks fue que se

tomara tan en serio los informes subjetivos de sus pacientes. Si ya era algo evidente en su libro de 1970 Migraña, quedaba aún más patente en su extraordinario Despertares, de 1973. Este segundo libro recogía con exquisito detalle los periplos clínicos de un grupo de pacientes crónicos «mudos acinéticos» que padecían encefalitis letárgica. Esta enfermedad también se conocía como «enfermedad del sueño», aunque los pacientes no estaban literalmente dormidos, sino que no mostraban iniciativa ni pulsión espontáneas. Sacks los «despertó» administrándoles levodopa, un fármaco que aumenta la disponibilidad de dopamina. Sin embargo, tras recuperar la capacidad de actuar, enseguida se volvieron demasiado impulsivos, maníacos y, finalmente, psicóticos. Luego leí Con una sola pierna, donde Sacks describía su propia experiencia subjetiva de una lesión del sistema nervioso. Poco después, en 1985, publicó El hombre que confundió a su mujer con un sombrero, una serie de estudios de casos que ofrecían una visión esclarecedora de los trastornos neuropsicológicos desde la perspectiva de ser un paciente neurológico. Este libro consagró la fama de Sacks.

Aquellos libros no tenían nada que ver con mis manuales de neuropsicología, que diseccionaban las funciones mentales como lo haríamos con las funciones de cualquier órgano del cuerpo. Aprendí, por ejemplo, que el lenguaje lo producía el área de Broca en el lóbulo frontal izquierdo, que la comprensión del habla tenía lugar en el área de Wernicke, unos centímetros más atrás, en el lóbulo temporal, y que la capacidad de repetir lo que te dicen estaba mediada por el fascículo arqueado, un conducto de fibras que conecta estas dos regiones. También aprendí que los recuerdos se codifican en el hipocampo, se almacenan en la neocorteza y se recuperan por mecanismos frontolímbicos.

¿De verdad el cerebro era tan equiparable al estómago y a los pulmones? Hay una diferencia obvia: existe «algo que es como» ser un cerebro, y esto no sucede con ninguna otra parte del cuerpo. Las sensaciones que ubicamos en otros órganos corporales no las sienten los propios órganos, sino que estos mandan impulsos nerviosos que solo se sienten cuando llegan al cerebro. Cabía pensar, pues, que esta propiedad tan característica del tejido cerebral —la capacidad de sentir, percibir y pensar cosas— existía por alguna razón. Esta propiedad parecía hacer algo. Y, si lo hacía —si la experiencia subjetiva tenía efectos causales sobre el comportamiento, como parece ocurrir cuando decidimos de pronto prepararnos un té—, sería un grave error omitirla de nuestras explicaciones científicas. Sin embargo, eso es justo lo que ocurría en la década de 1980. Mis profesores no hablaron en ningún momento de qué se siente al comprender el habla o recuperar un recuerdo, y aún menos de por qué siquiera se siente

algo.

Quienes sí tenían en cuenta la perspectiva subjetiva no eran tomados en serio por los neurocientíficos propiamente dichos. No sé hasta qué punto se sabe que las publicaciones de Sacks fueron muy ridiculizadas por sus colegas. Un articulista llegó a llamarlo «el hombre que confundió a sus pacientes con una carrera literaria». Y todo aquello apesadumbraba a Sacks. ¿Cómo se puede describir la vida interior de los seres humanos sin contar sus historias? También Freud se lamentaba un siglo antes al hablar de sus historiales clínicos:

Todavía me resulta singular que los historiales clínicos por mí escritos se lean como novelas breves, y de ellos esté ausente, por así decir, el sello de seriedad que lleva estampado lo científico. Debo consolarme con la reflexión de que la responsable de ese resultado es la naturaleza misma del asunto, y no una predilección mía.[4]

Sacks se mostró encantado cuando le envié esta cita.[5] En mi caso, la primera vez que leí esas líneas vi que no era el único que había llegado a la neuropsicología con la esperanza de que esta disciplina me ayudara a entender el modo en que el cerebro genera la subjetividad. Sin embargo, pronto te quitan esa idea de la cabeza, con la advertencia de que dedicarse a cuestiones tan inextricables es «malo para tu carrera», de ahí que la mayoría de los estudiantes de Neurociencia olviden poco a poco por qué eligieron esos estudios y acaben identificándose con el dogma del cognitivismo, que aborda el cerebro como algo no muy distinto a un teléfono móvil.

El único aspecto de la conciencia que era un tema científico respetable en la década de 1980 era el mecanismo cerebral de la vigilia por oposición al sueño. En otras palabras, el «nivel» de conciencia se consideraba un tema respetable, pero no su «contenido». Así pues, decidí centrar mi investigación doctoral en algún aspecto del sueño. En concreto, elegí estudiar el aspecto subjetivo del sueño, es decir, los mecanismos cerebrales de los sueños. Después de todo, la experiencia onírica no es sino una intrusión paradójica de la conciencia («vigilia») en el sueño. Para mi sorpresa, había una enorme laguna en la bibliografía sobre este tema: nadie había descrito de forma sistemática cómo afectaban a los sueños las lesiones en distintas partes del encéfalo. De modo que decidí hacerlo yo.

Esa naturaleza subjetiva de los sueños es precisamente el principal obstáculo para su estudio. En general, los fenómenos mentales solo tienen un testigo, el sujeto que los experimenta introspectivamente, que luego los comunica a otras personas de manera indirecta mediante las palabras. Pero los sueños lo ponen aún más difícil, porque solo se pueden relatar con carácter retrospectivo, cuando el sueño ya ha terminado y quien lo ha tenido se ha despertado. Todos sabemos lo poco fiable que es nuestra memoria en lo relativo a los sueños. ¿Qué clase de «datos» son esos?[6] Por ese motivo, a partir de mediados del siglo XX los sueños fueron un frente importante en la transición del conductismo a lo que más tarde se convertiría en la neurociencia cognitiva.

El electroencefalograma se aplicó por primera vez al estudio del sueño a principios de la década de 1950. Lo utilizaron dos neurofisiólogos, Eugene Aserinsky y Nathaniel Kleitman. Según su hipótesis inicial, el nivel de actividad cerebral disminuía al dormirnos y aumentaba al despertarnos, lo que los llevó a predecir que al dormirnos, aumentaba la amplitud de nuestras ondas cerebrales —uno de los elementos medidos por la electroencefalografía— y disminuía su frecuencia —el otro elemento medido—, mientras que al despertarnos ocurriría lo contrario (véase fig. 10, p. 151).

Cuando el cerebro desciende a lo que ahora se denomina «sueño de ondas lentas», vemos exactamente lo que predijeron Aserinsky y Kleitman, confirmando así su hipótesis. Lo que no esperaban era que unos noventa minutos después de haberse dormido la persona de la que se obtienen las grabaciones —y aproximadamente cada noventa minutos a partir de entonces, en ciclos regulares—, sus ondas cerebrales volvieran a acelerarse y casi alcanzaran los niveles de vigilia, aunque siguiera dormida.[7] Aserinsky y Kleitman denominaron a estos curiosos estados de activación cerebral «sueño paradójico» (la paradoja radica en que el cerebro está fisiológicamente despierto pese a estar profundamente dormido).

En este peculiar estado ocurren otras cosas. Los ojos se mueven muy deprisa —de ahí que el sueño paradójico pasara a denominarse «movimiento ocular rápido» o «sueño REM»—, pero el cuerpo por debajo del cuello está temporalmente paralizado. También se producen cambios autónomos drásticos, como la reducción del control de la temperatura corporal central y la turgencia de los genitales, que provoca erecciones visibles en los hombres. Lo que no se entiende es que la ciencia no se diera cuenta de todo esto hasta 1953.

Basándose en estas observaciones, Aserinsky y Kleitman formularon

otra hipótesis nada descabellada: que el sueño REM es la base fisiológica del estado psicológico que llamamos «soñar». En consecuencia, predijeron que los despertares de la fase de sueño REM conducirían a informes de sueños, pero no así los despertares de la fase de sueño de ondas lentas (no REM). Junto con William Dement —apellido desafortunado donde los haya—, pusieron a prueba esta predicción y la confirmaron: el porcentaje de despertares que daban lugar a informes de sueños era de casi un 80 por ciento en el sueño REM, frente a apenas un 10 por ciento en la fase de sueño no REM. A partir de ese momento, el sueño REM se consideró sinónimo del acto de soñar.[8] Excelente noticia, dado que, al disponer por fin de un marcador objetivo de los sueños, los neurocientíficos ya podían ejercer la ciencia seriamente sin tener que lidiar con las complicaciones metodológicas que les suponían los informes verbales, retrospectivos y de testigo único sobre experiencias subjetivas fugaces.

Había otra razón para celebrar el haberse librado de los sueños: el incómodo papel que estos habían desempeñado cuando apareció el psicoanálisis. A diferencia de las respuestas de la corriente dominante a la metafísica mente-cuerpo que caracterizó la ciencia mental en la segunda mitad del siglo XX, los psicoanalistas no tenían reparos en tratar los informes introspectivos como datos. De hecho, los informes obtenidos por «asociación libre» (muestreo no estructurado de la corriente de conciencia) eran los principales datos de la investigación psicoanalítica. Aplicando este método, Sigmund Freud llegó a la conclusión de que, a pesar de la apariencia absurda de las experiencias oníricas «manifiestas», su contenido «latente» (la historia subyacente, que él infirió de las asociaciones libres de la persona que soñaba) revelaba una función psicológica coherente, a saber, la realización de deseos.

Según Freud, soñar es lo que hacemos cuando las necesidades biológicas que generan el comportamiento de vigilia se liberan de su inhibición durante el sueño. Los sueños son intentos de satisfacer esas necesidades, que siguen con sus exigencias incluso cuando dormimos. Sin embargo, los sueños responden a esas necesidades de forma alucinatoria, lo que nos permite permanecer dormidos (en lugar de despertarnos para satisfacer realmente nuestras pulsiones). Como las alucinaciones son una característica central de las enfermedades mentales, Freud utilizó esta teoría en su trascendental obra La interpretación de los sueños (1900) para esbozar, con pinceladas gruesas, un modelo del funcionamiento general de la mente en la salud y en la enfermedad.

Como dijo Freud, «el psicoanálisis se basa en el análisis de los sueños».

[9] Pero ya hemos visto que resulta muy difícil estudiar empíricamente los sueños; de ahí que los conductistas los excluyeran de la ciencia. Por otra parte, el edificio teórico que Freud construyó sobre los sueños no era mejor que sus cimientos. El gran filósofo de la ciencia Karl Popper declaró que la teoría psicoanalítica era «pseudocientífica» porque no daba lugar a predicciones falsables mediante experimentos.[10] ¿Cómo falsar la hipótesis de Freud de que los sueños expresan los deseos latentes? Si no es necesario que esos deseos aparezcan en el sueño manifiesto (relatado), entonces cualquier sueño puede ser «interpretado» de forma que se adapte a los requisitos de la teoría. No es de extrañar, por tanto, que cuando el descubrimiento del sueño REM les permitió saltar del material efímero de los informes de los sueños a sus indicadores fisiológicos concretos, los neurocientíficos se libraran de los sueños en sí como quien suelta corriendo peces que se le escurrían de las manos.

El descubrimiento del sueño REM en la década de 1950 fue el pistoletazo de salida de una carrera para identificar su base neurológica, dado que la función del sueño REM podría revelar el mecanismo objetivo de los sueños, cuya elucidación daría a la psiquiatría de la época una base científica más respetable. (Facilitó mucho la investigación el hecho de que el sueño REM se dé en todos los mamíferos). Ganó la carrera Michel Jouvet, en 1965. En una serie de experimentos quirúrgicos con gatos, demostró que el sueño REM no lo generaba el prosencéfalo —que incluye la corteza, la parte superior del encéfalo, de tamaño tan impresionante en los seres humanos que se la considera el órgano de la mente—, sino el tronco encefálico, una estructura en principio mucho más humilde y que casi se remonta a los orígenes de la evolución.[11] Jouvet llegó a esta conclusión al observar que conforme practicaban cortes progresivos del encéfalo, de arriba abajo, la pérdida del sueño REM no se producía hasta que el corte había alcanzado el nivel de una «modesta» estructura del tronco encefálico conocida como «protuberancia» (véase fig. 1, p. 33).[12]

Fue Allan Hobson, alumno de Jouvet, quien acabó de concretar el tema. Hobson identificó con precisión qué conjuntos de neuronas pontinas generaban el sueño REM y, con él, los sueños. A mediados de la década de 1970 se evidenció que todo el ciclo sueño-vigilia —incluidos todos los fenómenos del sueño REM antes mencionados y las distintas fases del sueño no REM— lo orquestaba un número reducido de núcleos del tronco encefálico que interactuaban entre sí.[13] Los que controlan el sueño REM funcionan como un simple interruptor de encendido-apagado. Las neuronas que encienden el sueño REM se hallan en el tegmento mesopontino (véase fig. 1) y liberan una sustancia neuroquímica llamada acetilcolina por todo el prosencéfalo.

La acetilcolina provoca la excitación, esto es, aumenta el «nivel» de conciencia (por ejemplo, la nicotina la potencia, y eso ayuda a la persona a concentrarse). Las neuronas del tronco encefálico que apagan el sueño REM se encuentran a mayor profundidad en la protuberancia, en el rafe dorsal y el complejo del locus cerúleo (véase de nuevo fig. 1). Liberan serotonina y noradrenalina, respectivamente. Al igual que la acetilcolina, estas sustancias neuroquímicas modulan distintos aspectos del nivel de conciencia.

Combinando estos hallazgos con el hecho de que el sueño REM se apaga y enciende de forma automática más o menos cada noventa minutos, como un reloj, Hobson no tardó en sacar la inevitable conclusión: «La principal fuerza motivadora de los sueños no es psicológica, sino fisiológica, ya que el momento de aparición y la duración del sueño onírico son bastante constantes, lo que sugiere una génesis preprogramada y neuronalmente determinada».[14]

Puesto que el sueño REM surge del tronco encefálico colinérgico —una parte antigua y modesta del encéfalo alejada de la imponente corteza en la que se supone que tiene lugar toda la acción de la psicología humana—, Hobson añadió que los sueños no podían estar motivados por deseos, porque eran «motivacionalmente neutros».[15] Por lo tanto, según Hobson, la idea de Freud de que los sueños están determinados por deseos latentes tenía que ser un error. El significado que Freud atribuía a los sueños no era más intrínseco a ellos que a las manchas de tinta. Se proyecta en ellos, pero sin estar en ellos. Desde el punto de vista científico, la interpretación de los sueños no era mejor que leer las hojas de té.

Como todo el psicoanálisis se basaba en el método que Freud utilizó para estudiar los sueños, ahora ya se podía descartar en su integridad el cuerpo teórico que se derivó de ello. Después de que Hobson echase por tierra la idea de que los sueños podían significar algo, la psiquiatría pudo por fin abandonar su dependencia histórica de los informes introspectivos y basarse en métodos neurocientíficos objetivos (sobre todo neuroquímicos) de investigación y tratamiento. En consecuencia, mientras que en la década de 1950 era casi imposible llegar a ser profesor titular de Psiquiatría en ninguna de las grandes universidades estadounidenses si no eras psicoanalista, hoy día ocurre lo contrario: es casi imposible llegar a ser profesor de Psiquiatría si eres psicoanalista.

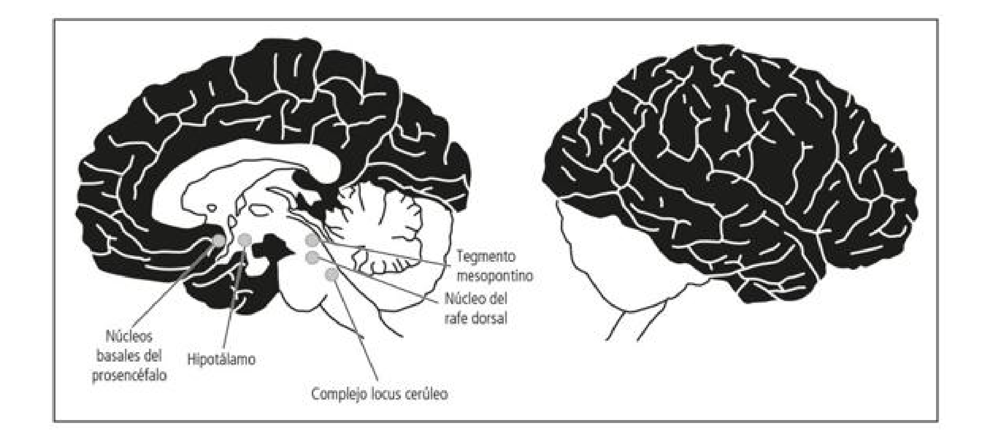


Figura 1

. La imagen de la izquierda es una vista medial del encéfalo (corte por el medio) y la de la derecha es una vista lateral (de perfil). La figura muestra la corteza (negra) y el tronco del encéfalo (blanco). Solo se indican los núcleos del tronco encefálico considerados importantes para el control del sueño REM: el tegmento mesopontino, el núcleo del rafe dorsal y el complejo locus cerúleo. También se muestra la ubicación de los núcleos basales del prosencéfalo (debajo de la corteza) y del hipotálamo, cuya relevancia se pone de manifiesto más adelante.

Nada de aquello me llamó especialmente la atención por aquel entonces. El tema central de mi investigación doctoral parecía bastante claro y no tenía nada que ver con las batallas entre los legados freudiano y conductista. Yo solo quería saber lo siguiente: ¿cómo afectaban a la experiencia real de soñar las lesiones sufridas en distintas partes del prosencéfalo y su corteza cerebral? A fin de cuentas, si todo pasaba en el prosencéfalo, psicológicamente hablando, algo debía de hacer el prosencéfalo con los sueños.

El Departamento de Neurocirugía de la Universidad de Witwatersrand tenía salas en dos hospitales universitarios: el Hospital Baragwanath y el Hospital General de Johannesburgo. Baragwanath era un enorme hospital, anteriormente militar, situado en el municipio «no europeo» de Soweto. En pleno apogeo del apartheid en Sudáfrica, aquel lugar era un mar de miseria humana. En cambio, el Hospital General de

Johannesburgo, reservado a los «europeos», era un hospital académico de vanguardia, un monumento a la desigualdad racial. El Departamento de Neurocirugía también tenía camas en la Unidad de Rehabilitación Cerebral y de la Columna Vertebral del Hospital General de Edenvale, ubicado en un antiguo edificio colonial de la zona residencial de Johannesburgo. A partir de 1985 ejercí en los tres centros, donde examiné a cientos de pacientes al año. A 361 de ellos los incluí en mi investigación doctoral, que se prolongó durante los cinco años siguientes.

Tras aprender a utilizar la electroencefalografía y otras tecnologías afines, y a reconocer las ondas cerebrales características asociadas a las distintas fases del sueño, supe cómo despertar a las personas durante la fase REM, cuando era más probable que estuvieran soñando. Asimismo, entrevisté a pie de cama a pacientes neurológicos, preguntándoles por los cambios producidos en sus sueños, y luego les hice un seguimiento de días, semanas y meses. La idea era investigar si el contenido de los sueños se veía sistemáticamente afectado por lesiones localizadas en distintas partes del encéfalo. Pese a la mala reputación de los informes sobre los sueños, supuse que si varios pacientes con lesiones en la misma zona del encéfalo señalaban el mismo cambio en el contenido de los sueños, había motivos para creerlos. Este método se llama «correlación clínico-anatómica»: mediante el estudio clínico de las capacidades psicológicas de los pacientes, se observa la alteración de alguna función mental a causa de la lesión de una parte del encéfalo; luego se relaciona esa alteración con la ubicación de la lesión, lo que lleva a descubrir pistas sobre la función de la estructura cerebral dañada, pistas que a su vez conducen a hipótesis comprobables. El método se había aplicado sistemáticamente unas décadas antes a todas las funciones cognitivas principales, como la percepción, la memoria y el lenguaje, pero nunca se había aplicado a los sueños.

Al principio, me daba un poco de reparo hablar con personas tan enfermas sobre sus sueños. Muchos tenían por delante una operación cerebral a vida o muerte o acababan de someterse a ella, y en esas circunstancias temía que consideraran frívolas mis preguntas. Sin embargo, para mi sorpresa, mis pacientes se mostraron más que dispuestos a describir los cambios que las enfermedades neurológicas habían provocado en su vida mental.

En la época en la que empecé mi investigación, se habían publicado varios casos clínicos en los que se demostraba que el mismo efecto observado en animales de experimentación se producía en seres humanos: que las lesiones en el tegmento mesopontino (véase fig. 1)

suprimían el sueño REM. Lo sorprendente, sin embargo, es que nadie se había molestado en indagar sobre los cambios en los sueños de esos pacientes. Este es el ejemplo más claro de los prejuicios de la neurociencia contra los datos subjetivos.[16]

En mi investigación, esperaba encontrar lo obvio: que los pacientes con lesiones en la corteza visual tuvieran sueños no visuales; que los pacientes con lesiones en la corteza del lenguaje tuvieran sueños no verbales; que los pacientes con lesiones en la corteza somatosensorial y motora tuvieran sueños hemipléjicos, y así sucesivamente: los principios básicos de la correlación cerebro-comportamiento. Ese era el vacío que quería llenar y que, afortunadamente, llené.[17]

Sin embargo, para mi asombro, junto a todas las cosas obvias que observé, descubrí también que los pacientes con lesiones en la parte del encéfalo que genera el sueño REM seguían teniendo sueños. Además, los pacientes en los que habían desaparecido los sueños tenían la lesión en una parte muy distinta del encéfalo. Así pues, los sueños y el sueño REM eran lo que llamamos fenómenos «doblemente disociables».[18] Estaban correlacionados entre sí (es decir, solían producirse al mismo tiempo), pero no eran lo mismo.[19]

Durante casi cincuenta años, en todo el campo de la ciencia del sueño, los investigadores del cerebro han confundido correlación con identidad. En cuanto determinaron que los sueños acompañaban a la fase REM concluyeron que ambos eran lo mismo y se deshicieron de la parte subjetiva de la correlación que tantos problemas daba. Luego, con muy pocas excepciones, estudiaron únicamente el sueño REM, sobre todo en animales de experimentación, que no pueden proporcionar informes introspectivos. El error no salió a la luz hasta que no empecé a llamar la atención de los neurocientíficos sobre la experiencia de los sueños en pacientes neurológicos.

Cuando, a principios de la década de 1990, informé por primera vez de la supresión de sueños a causa de lesiones sufridas en una parte del encéfalo distinta de la que genera el sueño REM, insistí mucho en que la zona decisiva no estaba en el tronco encefálico.[20] Quería hacer hincapié en la naturaleza mental de los sueños, y todos sabíamos que las funciones mentales residen en la corteza.

De hecho, descubrí dos zonas cuya lesión provocaba la supresión de los sueños pero preservaba el sueño REM. La primera estaba en la corteza, en el lóbulo parietal inferior (véase fig. 2, p. 38). El hallazgo no era sorprendente, dada la importancia del lóbulo parietal para la memoria a corto plazo. Si un paciente no puede retener el contenido

de su recuerdo en la reserva de datos de su vida consciente, ¿cómo puede tener un sueño? La segunda zona cerebral que descubrí fue mucho más interesante: la sustancia blanca del cuadrante ventromesial de los lóbulos frontales, que conecta la corteza frontal con varias estructuras subcorticales. Fue un hallazgo totalmente inesperado; no hay nada en las funciones de esta parte del encéfalo que tenga una relación clara con la experiencia manifiesta de soñar, y, sin embargo, debe de aportar algo crucial al proceso, porque su lesión provocó de forma fidedigna el cese total de los sueños.

Digo «de forma fidedigna» a pesar de que solo he informado de nueve casos de pérdida de los sueños entre mis pacientes con lesiones en el lóbulo frontal (y cuarenta y cuatro casos del tipo parietal). Estas lesiones son sumamente raras en la práctica clínica ordinaria, lo que no obsta para que la correlación sea fiable. En la primera mitad del siglo XX se hicieron millares de intervenciones quirúrgicas de la sustancia blanca frontal ventromesial mediante una técnica denominada «leucotomía prefrontal modificada». Era una época de desenfreno en la que los psiquiatras descubrieron que algunas enfermedades mentales graves se podían aliviar mediante la destrucción quirúrgica completa de los lóbulos prefrontales (técnicamente conocida como «lobotomía frontal»); pero también se dieron cuenta de que aquel procedimiento tan radical tenía muchos «efectos secundarios», por citar su propio eufemismo. Así pues, redujeron la extensión de la lesión y buscaron la parte más pequeña posible de los lóbulos frontales que se podía desconectar del resto del encéfalo para seguir obteniendo los resultados deseados. La solución fue el procedimiento modificado de Walter Freeman y James Watts. Requería la inserción de una diminuta cuchilla giratoria a través de las cuencas oculares, que cortaba la materia blanca en el cuadrante ventromesial de los lóbulos frontales (leucotomía prefrontal), en la ubicación exacta de la lesión de mis nueve pacientes.

Volví, pues, a la antigua bibliografía psicoquirúrgica para ver si podía confirmar lo que observaba en mis casos.[21] Tenía motivos para albergar esperanzas de que los médicos que examinaron a los pacientes de la leucotomía clásica les hubieran preguntado por los sueños tras las operaciones; al fin y al cabo, en aquella época los psiquiatras todavía se tomaban en serio los sueños. Mis esperanzas se vieron cumplidas: descubrieron que la leucotomía prefrontal tenía tres efectos psicológicos principales. En primer lugar, reducía los síntomas psicóticos positivos (alucinaciones y delirios); en segundo, reducía la motivación; y, en tercer lugar, provocaba la pérdida de los sueños. De hecho, uno de los primeros investigadores psicoquirúrgicos llegó a sugerir que la conservación de los sueños tras la operación era un mal

signo pronóstico.[22]

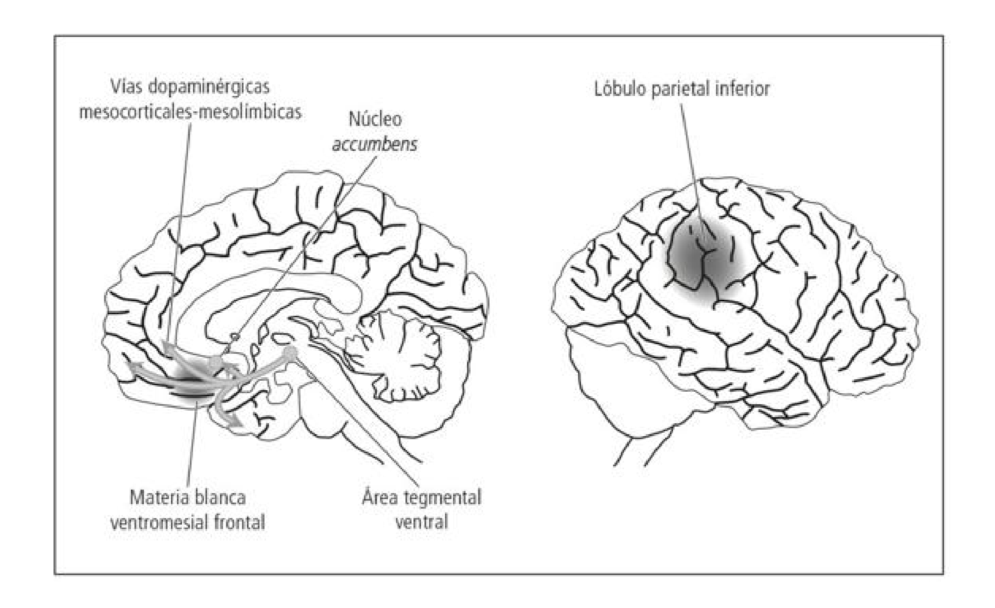


Figura 2

. Las dos áreas de lesión que conducen a la supresión de los sueños aparecen sombreadas en esta figura: la sustancia blanca ventromesial del lóbulo frontal (izquierda) y el lóbulo parietal inferior de la corteza (derecha). También se muestra el área tegmental ventral del tronco encefálico y las principales vías de fibras que parten de ella, a saber, las vías dopaminérgicas mesocorticales-mesolímbicas. Nota: la lesión en el lóbulo frontal ventromesial afecta a estas vías subcorticales, que discurren por debajo de la corteza, no dentro de ella. Un destino importante de estas vías es el núcleo accumbens, también mostrado en la figura.

Este último punto me ayudó a conjeturar a cuál de los numerosos circuitos neuronales del cuadrante ventromesial de los lóbulos frontales era más probable que se debiera la pérdida de los sueños. También me dio una primera pista de por qué debíamos buscar a nuestro culpable en esa inesperada región del encéfalo. ¿Qué son los sueños sino alucinaciones y delirios? Por eso sería un mal signo pronóstico si persistieran tras la leucotomía.

En realidad, el tratamiento neuroquirúrgico de las alucinaciones y los

delirios no se abandonó por motivos éticos, sino que cayó en desuso cuando se hizo evidente que podían obtenerse resultados terapéuticos equivalentes con menos morbilidad y mortalidad recurriendo a unos fármacos que empezaron a distribuirse ampliamente en la década de 1950: los «tranquilizantes mayores». Lo que hacían estos fármacos —y lo que siguen haciendo los «antipsicóticos» modernos— era bloquear la dopamina neuroquímica en los terminales de un circuito cerebral conocido como «sistema dopaminérgico mesocortical-mesolímbico» (véase fig. 2). Dado que este circuito se corta con la leucotomía prefrontal, como ocurrió en mis nueve pacientes con lesiones de origen natural, formulé la hipótesis de que ese podía ser el sistema que genera los sueños.

Otros experimentos confirmaron mi hipótesis. Ya se había visto que la estimulación farmacológica del circuito mencionado aumentaba la frecuencia, duración e intensidad de los sueños, y sin efectos proporcionales sobre el sueño REM.[23] El fármaco en cuestión era la levodopa, el mismo que Oliver Sacks había utilizado para «despertar» a sus pacientes posencefalíticos. Hace tiempo que los neurólogos que utilizan estimulantes dopaminérgicos para el tratamiento de la enfermedad de Parkinson saben el cuidado que hay que tener para no llevar a sus pacientes a la psicosis, como le pasó a Sacks; la aparición de sueños más vívidos de lo normal suele ser el primer indicio de dicho efecto secundario.[24] Hubo observaciones posteriores de crucial importancia; por ejemplo, que las neuronas que constituyen este circuito (cuyas somas se encuentran en el área tegmental ventral) se disparan a tasas máximas durante el sueño onírico[25] y, al mismo tiempo, suministran cantidades máximas de dopamina a sus dianas en el núcleo accumbens (véase fig. 2).[26] Por eso ahora se acepta sin mayor discusión que se puede soñar con independencia del sueño REM y que el circuito dopaminérgico mesocortical-mesolímbico es, de hecho, el principal impulsor de los sueños.[27]

La lesión de las vías colinérgicas en el cuadrante ventromesial de los lóbulos frontales —que surgen de los núcleos basales del prosencéfalo (véase fig. 1)— produce el efecto contrario al de las lesiones de las vías dopaminérgicas: más sueños en lugar de menos. Hobson había afirmado que la acetilcolina era el generador motivacionalmente neutro de los sueños, pero el resultado es el mismo cuando se bloquea la acetilcolina con fármacos que cuando se lesionan sus vías. Ahora se sabe que los fármacos anticolinérgicos —bloqueantes de la acetilcolina — provocan demasiados sueños.[28] Así pues, el bloqueo del sistema nervioso que según Hobson era responsable de los sueños tiene el efecto contrario al que predijo su teoría.

Enseguida se vio que la neurociencia le debía una disculpa a Freud. Porque si hay una parte del encéfalo que puede considerarse responsable de los «deseos», es sin duda el circuito mesocortical-mesolímbico de la dopamina, que es todo menos motivacionalmente neutro. Edmund Rolls (y muchos otros) lo denomina «sistema de recompensa» del cerebro.[29] Kent Berridge lo llama «sistema de querer». Jaak Panksepp lo denomina «sistema de BÚSQUEDA» y destaca su papel en la función de búsqueda de alimento.[30] Se trata del circuito cerebral responsable de «los comportamientos de exploración y búsqueda más motivados que un animal es capaz de mostrar».[31] También es el circuito que impulsa los sueños.[32]

Todo aquello no le hizo ninguna gracia a Hobson. Me invitó a presentar mis conclusiones ante su grupo de investigación del Departamento de Neurofisiología de Harvard. Al principio las aceptó e incluso publicó una reseña favorable del libro que escribí sobre el tema en 1997, en la que señalaba que mis hallazgos clínico-anatómicos se veían confirmados hasta el último detalle por los estudios de neuroimagen de Allen Braun (véase fig. 3, p. 52).[33] Sin embargo, cuando se dio cuenta de que aquellos avances podrían reivindicar una consideración bastante freudiana de los sueños, me escribió diciéndome que estaba dispuesto a respaldar públicamente mis hallazgos a condición de que yo no sostuviera que confirmaban a Freud. Hasta aquí la supuesta objetividad de la neuropsicología.

No obstante, mi descubrimiento albergaba otro elemento muy sorprendente. La primera vez que di con él no presté demasiada atención al hecho de que las neuronas que impulsaban este circuito se hallasen en el tronco encefálico (como las de los circuitos que generan el sueño REM). Como ya he dicho, quería hacer hincapié en la naturaleza mental del sueño. Allen Braun, el especialista en neuroimagen al que acabo de referirme, no pudo más que señalarme amablemente mi descuido. En el contexto de la discrepancia científica entre Hobson y yo sobre los circuitos cerebrales que impulsan el proceso del sueño (dopaminérgicos o colinérgicos), Braun escribió:

Lo curioso es que, tras defender el posible papel fundamental de las estructuras del prosencéfalo en el sistema onírico, Solms acaba sugiriendo que son los aferentes dopaminérgicos a estas regiones los que [generan los sueños], con lo que vuelve a situar al instigador de los sueños en el tronco encefálico.[34]

Braun concluyó: «Diría que estos caballeros están llegando a un terreno común».[35] En los años noventa, al igual que el resto de la neuropsicología, yo pensaba que toda la acción psicológica se desarrollaba en la corteza, y por ese motivo me centré en el hecho de que las vías de sustancia blanca que me interesaban estaban en los lóbulos frontales, el lugar donde se localizaban las lesiones de mis nueve casos. Sin embargo, todos los núcleos centrales del tronco encefálico envían largos axones hacia arriba, al prosencéfalo (véase fig. 2). Las somas de estas neuronas se hallan en el tronco encefálico, aunque las fibras que salen de ellas (los axones) terminan en la corteza, lo cual respalda la principal función de excitación de estos núcleos del tronco encefálico, conocidos en su conjunto como «sistema reticular activador». Eran precisamente estas vías activadoras las que estaban lesionadas en mis nueve pacientes y en los cientos de casos documentados que los precedieron sobre pacientes sometidos a leucotomía que no soñaban.

A partir de 1999, empujado en parte por los comentarios de Braun sobre las implicaciones de mi descubrimiento, centré mi atención en los demás sistemas de excitación del tronco encefálico. El trabajo más interesante en este campo lo estaba haciendo Jaak Panksepp, cuyo libro enciclopédico Neurociencia afectiva (1998) recogía con exquisito detalle todo tipo de pruebas que respaldaban su idea de que estos sistemas supuestamente mecánicos, responsables de regular solo el «nivel» de conciencia, generaban un «contenido» propio.

Aquel cambio de rumbo acabaría teniendo mucha importancia.

[1] Popper, 1963. Lo cierto es que no todo el mundo está de acuerdo con esta formulación. Sin embargo, es la que defienden casi todos los científicos naturales. (Ahora que están leyendo estas notas, conviene una aclaración sobre ellas: están pensadas sobre todo para lectores académicos que tengan interés [o formación previa] en la bibliografía técnica de la que bebe este libro. Los lectores generales, que son mi público principal, pueden pasarlas perfectamente por alto).

[2] Freud, 1893a, p. 13. [La traducción de las citas es nuestra, salvo que se mencione lo contrario, para no apartarnos de las argumentaciones del autor. No obstante, en la medida de la disponibilidad hemos consultado las traducciones citadas en la

bibliografía. (N. de la T.)].

[3] Oliver Sacks, Con una sola pierna, Barcelona: Anagrama, 1998, trad. de José Manuel Álvarez Flórez.

[4] Freud, 1895, p. 160. Unos treinta y seis años más tarde, Freud escribió una conmovedora carta a Albert Einstein sobre la falta de prestigio científico de la psicología en comparación con la física (Freud, 1994, p. 239): «En cualquier caso no es un asunto a lamentar el haber optado por la psicología. No hay tema más grande, más rico y más misterioso, digno de todos los esfuerzos del intelecto humano, que la vida de la mente. Sin duda la psicología es la más bella de todas las damas nobles; el único problema es que su caballero está condenado a amarla sin esperanzas».

[5] «Me encanta esa cita de Freud y me alegra mucho que la hayas localizado. Como señalas tú tan generosamente, podría decirse algo parecido de mis historias de casos y de las historias de casos neurológicos (al menos neuropsicológicos) en general. Yo lo he citado (aunque no sé si sobrevivirá, porque mi manuscrito es ya demasiado largo y contiene demasiadas notas al pie) en un texto general que acabo de terminar, "Scotoma", sobre el olvido y el abandono en la ciencia». (Carta de Sacks del 2 de enero de 1995).

[6] Como escribió el neurocientífico Semir Zeki en aquella época: «La mayoría [de nosotros] se encogería de espanto solo de pensar en investigar lo que parece un problema tan impenetrable». (Zeki, 1993, p. 343)

[7] Aserinsky y Kleitman, 1953.

[8] Dement y Kleitman, 1957.

[9] Véase Freud, 1912, p. 259: «Existe un producto psíquico que encontramos en las personas más normales y que, sin embargo, ofrece una singularísima analogía con los productos más extraños e intensos de la locura y que no ha sido para los filósofos más inteligible que la propia locura. Me refiero a los sueños. El psicoanálisis se basa en el análisis de los sueños; la interpretación de los sueños es la labor más completa que la joven ciencia ha llevado a cabo hasta hoy».

[10] Popper, 1963.

[11] Desde el punto de vista estrictamente anatómico, el tálamo no se considera parte del tronco encefálico. Sin embargo, dado que, en términos fisiológicos, algunos de sus núcleos «inespecíficos» forman

parte del sistema reticular activador, se agrupan con las funciones del tronco encefálico (de ahí la denominación «sistema de activación reticular-talámico extendido», ERTAS). Los núcleos talámicos «específicos», que actúan sobre todo como estaciones retransmisoras de señales sensoriales, se agrupan con las funciones de la corteza. En este libro, utilizaré «tronco encefálico» y «corteza» sobre todo para designar la división funcional-anatómica entre la excitación del ERTAS y la representación talamocortical, respectivamente. Por lo tanto, consideraré como estructuras del tronco encefálico no solo el tálamo y el hipotálamo «inespecíficos», sino también los núcleos basales del prosencéfalo. Para una visión contemporánea del sistema activador retículo-talámico ampliado, véase Edlow et al., 2012.

[12] Jouvet, 1965.

[13] Hobson, McCarley y Wyzinski, 1975.

[14] McCarley y Hobson, 1977, p. 1346.

[15] Ibid., p. 1219.

[16] El único paciente que informó de pérdida de los sueños posiblemente había sufrido lesiones más allá del tegmento mesopontino generador de REM (a causa de una hemorragia subaracnoidea traumática; Lavie et al., 1984), lo que hacía difícil asociar los cambios en sus sueños a una región cerebral concreta.

[17] Para una descripción completa de mis hallazgos, véase Solms (1997a). Presenté mi tesis en 1991, pero no la publiqué hasta seis años después.

[18] Solms, 2000a. El principio de la «doble disociación» en neuropsicología nos permite separar las funciones mentales por sus junturas naturales: si una lesión en el área X (del encéfalo) causa la pérdida de la función A pero no de la función B, y una lesión en el área Y causa la pérdida de la función B pero no de la función A, entonces la función A y la B no pueden ser la misma cosa. En otras palabras, en este caso la función del sueño REM y la función de los sueños no pueden ser la misma. Están correlacionadas (es decir, ocurren al mismo tiempo), pero no son lo mismo.

[19] Aparte de mis hallazgos sobre las lesiones, son muchos los datos que apoyan esta conclusión. Por ejemplo, hay un 50 por ciento de probabilidades de obtener informes de sueños durante los primeros minutos de sueño (en el estadio 2 descendente), mucho antes del primer episodio REM. Asimismo, los sueños que son completamente

indistinguibles de los REM se producen en el sueño no REM con una frecuencia cada vez mayor durante la fase matutina ascendente del ritmo diurno. Es lo que se denomina «efecto del final de la mañana». Por otra parte, aunque los sueños son mucho más frecuentes en el sueño REM que en el no REM, el sueño no REM es más frecuente que el sueño REM, lo que hace que al menos una cuarta parte de todos los sueños se produzcan durante el sueño no REM. Más información en Solms, 2000a.

[20] Solms, 1991, 1995.

[21] Frank, 1946, 1950; Partridge, 1950.

[22] Schindler, 1953.

[23] Hartmann et al., 1980.

[24] Sharf et al., 1978. Estudios posteriores demostraron que los antagonistas de la dopamina tienen el efecto contrario (Yu, 2007).

[25] Dahan et al., 2007.

[26] Lena et al., 2005.

[27] Solms, 2011.

[28] Solms, 2001.

[29] Es una mala denominación que nos retrotrae a la época conductista. Hay muchas variedades de «recompensa» (es decir, placer) en el cerebro.

[30] Rolls, 2014; Berridge, 2003; Panksepp, 1998.

[31] Panksepp, 1998, p. 155.

[32] La actividad de BÚSQUEDA dopaminérgica (a diferencia de otra actividad monoamínica) continúa con el inicio del sueño y llega a su pico durante el sueño REM. Quizá no sea casualidad que esto coincida con momentos de sacadas oculares rápidas. Los movimientos oculares en los humanos, al igual que el olfateo y el movimiento de los bigotes en los roedores, son un buen indicador de la activación de BÚSQUEDA (véase Panksepp, 1998).

[33] Pace-Schott y Hobson, 1998.

[34] Braun, 1999, p. 196.

[35] Ibid., p. 201.

02

Antes y después de Freud

En 1987 tomé otra decisión que me alejó del resto de mis colegas: decidí formarme como psicoanalista.[36] Mis incipientes hallazgos en la investigación de los sueños me habían convencido del papel fundamental que tenían los informes subjetivos en la neuropsicología y de que la oposición de mis compañeros a Freud los había llevado al error en más de un sentido. No obstante, lo que me acabó de decidir no fueron los resultados de mi investigación.

Lo que me convenció fue un seminario al que asistí en la Universidad de Witwatersrand, a mediados de los años ochenta, dirigido por un profesor de Literatura Comparada llamado Jean-Pierre de la Porte y que versaba sobre La interpretación de los sueños. Mi investigación doctoral me había despertado la curiosidad por esa obra de Freud. Como todo el mundo en aquella época, yo era escéptico con respecto al pensador austríaco. Durante la carrera se me había dicho que el psicoanálisis era una «pseudociencia». En las ciencias duras ya nadie se tomaba en serio a Freud —y quizá por eso el seminario se celebró en un departamento de Humanidades—, pero yo decidí asistir por la disposición de Freud a hablar sobre el contenido de los sueños, el tema de mi investigación.

Según De la Porte, las conclusiones teóricas a las que llegó Freud no se podían entender si antes no se había leído y asimilado un manuscrito suyo previo, de 1895, pero publicado póstumamente en los años cincuenta. El manuscrito se titulaba «Proyecto para una psicología científica»,[37] y en él Freud intentaba cimentar sobre una base neurocientífica sus primeras ideas sobre la mente.

Con ello seguía los pasos de su gran maestro, el fisiólogo Ernst von Brücke, miembro fundador de la Sociedad Física de Berlín. En 1842, Emil du Bois-Reymond formuló la misión de la Sociedad como sigue:

Brücke y yo hicimos un juramento solemne para poner en práctica esta verdad: «Las únicas fuerzas que están activas en el organismo son las fuerzas físicas y químicas comunes. Para explicar lo que actualmente dichas fuerzas no pueden explicar hay que encontrar la manera o forma específica de su acción por medio del método físico-

matemático o bien suponer la existencia de otras fuerzas tan dignas como las fuerzas químico-físicas inherentes a la materia, reducibles a las fuerzas de atracción y repulsión».[38]

Johannes Müller, apreciado maestro de los anteriores, se había preguntado cómo y por qué la vida orgánica difiere de la materia inorgánica. Llegó a la conclusión de que «los organismos vivos son esencialmente diferentes de las entidades no vivas porque contienen algún elemento no físico o se rigen por principios distintos a los de las cosas inanimadas».[39] En resumen, para Müller, los organismos vivos poseen una «energía vital» o «fuerza vital» que las leyes fisiológicas no pueden explicar. Según él, los seres vivos no pueden reducirse a los mecanismos fisiológicos que los componen porque son entes indivisibles con objetivos y propósitos, lo que atribuía al hecho de que poseen alma. Teniendo en cuenta que la palabra alemana Seele puede traducirse como «alma», pero también como «mente»,[40] el desacuerdo entre Müller y sus alumnos se parece mucho al actual debate entre filósofos como Thomas Nagel y Daniel Dennett sobre si la conciencia puede reducirse a leyes físicas (Nagel lo niega, Dennett lo afirma).

Lo que me sorprendió durante el seminario de De la Porte fue enterarme de que Freud —el investigador pionero de la subjetividad humana— no se había alineado con el vitalismo de Müller, sino más bien con el fisicalismo de Brücke. Así, en las primeras líneas de su «Proyecto» de 1895, escribió: «La intención es estructurar una psicología que sea una ciencia natural: es decir, representar los procesos psíquicos como estados cuantitativamente determinados de partículas materiales especificables».[41] Yo desconocía la formación neurocientífica de Freud, y solo después supe que, aunque le costó, abandonó los métodos de investigación neurológicos cuando vio claramente, en algún momento entre 1895 y 1900, que los métodos entonces disponibles no tenían capacidad para revelar la base fisiológica de la mente.

Sin embargo, para Freud fue un cambio de dirección que le compensó con creces, porque le obligó a examinar con mayor minuciosidad los fenómenos psicológicos per se y a dilucidar los mecanismos funcionales que los sustentaban. Todo ello dio lugar al método de investigación psicológica que acabó denominando «psicoanálisis». Su hipótesis fundamental era que los fenómenos subjetivos manifiestos (ahora llamados «explícitos» o «declarativos») tienen causas latentes (ahora llamadas «implícitas» o «no declarativas»). Es decir, Freud

sostenía que el hilo errático de nuestros pensamientos conscientes solo puede explicarse si suponemos asociaciones intermedias implícitas de las que no somos conscientes, idea que derivó en el concepto de las funciones mentales latentes y, a su vez, en la famosa conjetura de Freud sobre la intencionalidad «inconsciente».

Como a principios del siglo XIX no había métodos para investigar la fisiología de los fenómenos mentales inconscientes, la única forma de inferir sus mecanismos era la observación clínica. Lo que Freud aprendió con ella dio lugar a su segunda afirmación fundamental. Observó que los pacientes adoptaban una actitud nada indiferente respecto a las intenciones inconscientes que se les infería; parecía más una cuestión de no querer verlas que de no poder verlas. Freud recurrió a varias palabras para describir esa tendencia —resistencia, censura, defensa y represión—, señalando que evitaba la angustia emocional. Esto sirvió a su vez para revelar el papel crucial de los sentimientos en la vida mental y hasta qué punto son la causa de todo tipo de sesgos interesados. Aquellos hallazgos (ahora obvios) mostraron a Freud que algunas de las principales fuerzas motivadoras de la vida mental son totalmente subjetivas, pero también inconscientes. La investigación sistemática de esas fuerzas lo llevó a su tercera afirmación fundamental: la conclusión de que en última instancia lo que apuntalaba los sentimientos eran las necesidades corporales; de que la vida mental humana, no menos que la de los animales, estaba impulsada por los imperativos biológicos de supervivencia y reproducción. Para Freud, dichos imperativos constituían el vínculo entre la mente sintiente y el cuerpo físico.

Así y todo, adoptó un abordaje muy sutil de esa relación mente-cuerpo, pues vio que los fenómenos psicológicos que estudiaba no eran directamente reducibles a los fenómenos fisiológicos. Ya en 1891 había afirmado que no era posible atribuir los síntomas psicológicos a procesos neurofisiológicos sin antes reducir los fenómenos psicológicos y fisiológicos (las dos partes de la ecuación) a sus respectivas funciones subyacentes. Como ya he señalado antes, al hablar del procesamiento de la información, las funciones pueden realizarse en distintos sustratos.[42] Y según Freud, solo en el terreno común de la función podían reconciliarse la psicología y la fisiología. Su objetivo era explicar los fenómenos psicológicos mediante leyes funcionales «metapsicológicas» (esto es, «más allá de la psicología»).[43] Al intento de saltarse este nivel funcional de análisis, pasando directamente de la psicología a la fisiología, se lo conoce hoy día como la «falacia localizacionista».[44]

Queda claro que para Freud, cuando no para sus seguidores, el

psicoanálisis estaba pensado como una fase intermedia. Por mucho que desde el principio hubiese pretendido discernir las leyes que sustentan nuestra rica vida interior de experiencia subjetiva, para él la vida mental seguía siendo un problema biológico.[45] En 1914 escribió: «Es de prever que todas nuestras ideas provisionales en psicología se sostendrán algún día sobre unos cimientos orgánicos».[46] Freud anticipó con entusiasmo el día en que el psicoanálisis regresaría a su unión con la neurociencia:

La biología […] es realmente un dominio de infinitas posibilidades. Debemos esperar de ella la información más sorprendente y no podemos adivinar qué respuesta dará, dentro de algunos decenios […]. Quizá sean dichas respuestas tales que echen por tierra nuestro artificial edificio de hipótesis.[47]

Aquel no era el Freud tan peligrosamente especulativo del que me habían hablado en la universidad. Para mí, el «Proyecto» fue una revelación, tanto como lo había sido para el propio Freud, que por aquel entonces le escribió a su amigo Wilhelm Fliess:

En el transcurso de una noche ajetreada […] se levantaron de repente las barreras, cayeron los velos y fue posible ver desde los detalles de las neurosis hasta los determinantes de la conciencia. Todo parecía encajar, los engranajes estaban bien colocados; daba la impresión de que era realmente una máquina y que pronto funcionaría sola.[48]

Sin embargo, la euforia duró poco. Un mes después, Freud escribió: «Ya no puedo entender qué pensaba cuando urdí la “Psicología”; no puedo entender cómo llegué a infligírsela a mis lectores».[49] Al no contar con los métodos neurocientíficos apropiados, Freud se basó en «figuraciones, transposiciones y conjeturas» para traducir sus deducciones clínicas en términos primero funcionales y luego fisiológicos y anatómicos.[50] Tras un último intento de revisión (contenido en una larga carta que envió a Fliess el 1 de enero de 1896), se le perdió la pista al «Proyecto», hasta su reaparición unos cincuenta años más tarde. Con todo, las ideas que contenía —el «fantasma oculto», según James Strachey, el traductor de Freud al inglés— impregnaron toda su teorización psicoanalítica… a la espera

de futuros avances científicos.[51]

Hay dos ideas contenidas en el «Proyecto» que destacan ahora a la luz de los descubrimientos contemporáneos. La primera es que el prosencéfalo es un «ganglio simpático» que controla y regula las necesidades del cuerpo. La segunda es que estas necesidades son la fuerza que impulsa la vida mental, «el impulso primario del mecanismo psíquico».[52] Al no tener una comprensión neurobiológica de cómo se regulan estas necesidades corporales en el encéfalo —y mucho menos de cómo podrían explicarse «mediante el método físico-matemático»—, Freud no tuvo más remedio que «suponer la existencia de otras fuerzas tan dignas como las fuerzas químico-físicas inherentes a la materia», si quería mantenerse fiel a los ideales de la Sociedad Física de Berlín. Eran las que él llamaba fuerzas «metapsicológicas», las fuerzas que subyacen tras los fenómenos psicológicos, y aclaró que quería «transformar la metafísica en metapsicología».[53] Freud quería sustituir la filosofía por la ciencia, una ciencia de la subjetividad. Nos pidió que no juzgáramos con demasiada severidad sus deducciones especulativas sobre los procesos mentales latentes:

Esto se debe solo a que estamos obligados a trabajar con los términos científicos; esto es, con el idioma figurado propio de la psicología (o, más exactamente, de la psicología de la profundidad). Si no, no podríamos descubrir los procesos correspondientes; ni siquiera los habríamos percibido. Los defectos de nuestra descripción desaparecerían con seguridad si estuviéramos ya en posición de reemplazar los términos psicológicos por términos fisiológicos o químicos.[54]

Una de las nuevas fuerzas que Freud se vio obligado a inferir fue el concepto de «pulsión», que definió como «el representante psíquico de los estímulos que se originan en el interior del organismo y llegan a la mente, como medición de la demanda de trabajo que se hace a la mente a consecuencia de su conexión con el cuerpo».[55]

El concepto de «pulsión» de Freud —que él consideraba la fuente de toda «energía psíquica»— no difería mucho de la «energía vital» de Müller, pero estaba conectado a las necesidades corporales. Freud describió los mecanismos causales por los que las pulsiones se convierten en cognición intencional como una «economía de la fuerza

nerviosa».[56] Aun así, admitió sin ambages que era «totalmente incapaz de formarse una idea» de cómo las necesidades corporales pueden convertirse en una energía mental.[57]

Cuando leí aquellas palabras casi un siglo después, me di cuenta de que había llegado el momento de «reemplazar los términos psicológicos por términos fisiológicos o químicos», y de que nos correspondía a nosotros hacerlo. Por ejemplo, la fuerza impulsora de los sueños, que era «latente» en los informes subjetivos de los pacientes de Freud y cuya existencia se consideraba por ello mismo infalsable, quedaba claramente «manifiesta» en la evidencia objetiva obtenida con los métodos fisiológicos in vivo que no estaban disponibles en la época de Freud. Las imágenes de la figura 3 (p. 52), por ejemplo, producidas con tomografía por emisión de positrones, [58] muestran con claridad que el circuito de BÚSQUEDA «anhelante» se ilumina como un árbol de Navidad durante el sueño onírico, mientras que los lóbulos prefrontales inhibidores están esencialmente apagados. A partir de estos hallazgos, cuando nos invitaron a Hobson y a mí a debatir la credibilidad científica de la teoría freudiana de los sueños en la conferencia Science of Consciousness de 2006, nuestros colegas allí reunidos votaron dos a uno a favor de reinstaurar la viabilidad de la teoría.[59]

A pesar de todos sus defectos, el «yo» subjetivo nunca fue excluido del psicoanálisis, donde, por muy incómodo que le resultara al resto de la ciencia, ocupaba un lugar de honor. Muchos colegas científicos me aconsejaron que no asociara mi trabajo al psicoanálisis, dada su histórica mala reputación. Me decían que era como si un astrónomo se dejara asociar con la astrología. Pero a mí me parecía muy poco ético intelectualmente no reconocer a Freud por lo que merecía ser reconocido. Con independencia de en qué medida los logró, sus objetivos eran los correctos para una ciencia de la mente. Por ello llamé a mi enfoque «neuropsicoanálisis». Como he comentado, la neuropsicología que me enseñaron bien podría haberse llamado «neuroconductismo», dada su actitud hacia la subjetividad. Quería dejar claro que la neuropsicología que yo estaba desarrollando giraba en torno a la experiencia vivida. Con ese ánimo, tras escribir un artículo programático sobre la relación entre psicoanálisis y neurociencia, me puse manos a la obra.[60]

En 1989 me mudé a Londres para recibir formación psicoanalítica. A fin de poder seguir con mi labor investigadora y clínica, también acepté el cargo de profesor honorario de Neurocirugía en la Facultad

de Medicina del Royal London Hospital. Me encantó poder formar parte de la gran tradición neurológica de aquel centro, donde a mediados del siglo XIX había ejercido de médico John Hughlings Jackson, el padre fundador de la neurología y la neuropsicología británicas. El Royal London Hospital se hallaba en Whitechapel, una zona que durante siglos ha sido un imán para los inmigrantes y que, por tanto, siempre ha atendido a comunidades vulnerables. Me recordaba al Hospital Baragwanath de Soweto. Me sentía como en casa lejos de casa.

A principios de la década de 1990, un colega neurocirujano de Sudáfrica me remitió un paciente suyo, el señor S. Lo había operado diez meses antes para extirparle un tumor que, al crecer bajo los lóbulos frontales del cerebro, le estaba desplazando los nervios ópticos. Durante la intervención, el señor S. sufrió una pequeña hemorragia que interrumpió el riego sanguíneo del prosencéfalo basal (véase fig. 1). Los núcleos basales del prosencéfalo transmiten acetilcolina a varias estructuras corticales y subcorticales implicadas en la recuperación de los recuerdos a largo plazo. Se cree que estas vías colinérgicas interactúan con las vías dopaminérgicas (véase fig. 2), que forman el llamado «sistema de recompensa» que activa los comportamientos de «búsqueda», no solo en relación con las acciones físicas del mundo exterior, sino también con el mundo interior de las representaciones, las acciones imaginarias que surgen en el pensamiento y en los sueños.[61] A causa de la hemorragia, cuando el señor S. despertó de la intervención, sufría un profundo síndrome amnésico, conocido como «psicosis de Korsakoff», cuya característica principal es un estado onírico denominado «fabulación». Su memoria para los acontecimientos recientes había quedado tan desordenada que todo el tiempo recuperaba recuerdos falsos. Este déficit de búsqueda ya es incapacitante de por sí, pero en la amnesia fabuladora se ve agravado por el hecho de que los pacientes no controlan la fiabilidad de los recuerdos que recuperan falsamente, por lo que los tratan como si fueran verdaderos cuando es obvio que no lo son.

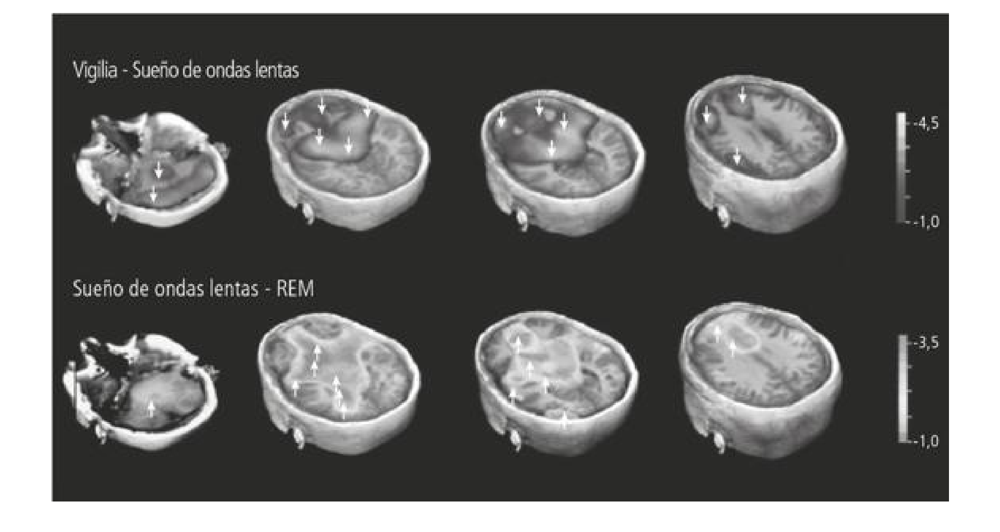


Figura 3

. Las filas horizontales muestran secciones transversales del encéfalo cada vez más altas (de izquierda a derecha). La superior muestra la diferencia entre el cerebro despierto y el cerebro dormido, donde el área sombreada representa la disminución de la activación cortical con el inicio del sueño. La fila inferior muestra la diferencia entre el sueño REM y el sueño no REM (ondas lentas), donde el área resaltada representa el aumento de la activación subcortical con el inicio del REM. La zona de mayor activación es donde se encuentra el sistema de BÚSQUEDA.

Por ejemplo, el señor S. creía que estaba en Johannesburgo, su ciudad natal, cuando en realidad acababa de llegar a Londres para consultarme. No recordaba el viaje, pero cuando lo corregí, insistió en que no podía estar en Londres. Entonces le pedí que mirase por la ventana, porque estaba nevando y en Johannesburgo no nieva. Pese a su sorpresa inicial, se recompuso y replicó: «No, no. Yo sé que estoy en Johannesburgo. Que uno esté comiendo pizza no significa que esté en Italia». El señor S. era ingeniero eléctrico y tenía cincuenta y seis años. Lo veía en mi consulta externa diaria seis veces por semana; el objetivo era orientarlo y ayudarlo a comprender cómo le fallaba la memoria. Pese a que la visita se repetía cada día a la misma hora y en el mismo lugar, nunca me reconocía de una sesión a otra como su terapeuta. Me reconocía la cara, pero siempre confundiéndome con otra persona a la que conocía en un contexto diferente, en la mayoría de los casos un colega ingeniero con el que intentaba resolver algún

problema electrónico o un cliente que buscaba su ayuda profesional. Dicho de otro modo, el señor S. me trataba como si yo necesitara su ayuda y no al revés. Otro de sus equívocos más frecuentes consistía en afirmar que ambos éramos estudiantes universitarios que tomaban algo juntos después de una actividad deportiva (una carrera de remos o un partido de rugby). Yo entonces era lo bastante joven para que la idea fuera creíble, pero el señor S. había sido estudiante hacía más de treinta años.

Tras cada sesión clínica, me reunía con la esposa del señor S. para contextualizar sus recuerdos erróneos e intentar esclarecer su significado. Esa era la principal diferencia entre mi planteamiento y el enfoque más tradicional que mis colegas daban a la «rehabilitación cognitiva». Frente a la preocupación convencional de los neuropsicólogos por el grado de trastorno de la memoria, medido desde el punto de vista de un tercero, a mí me interesaba más el contenido subjetivo de los errores del señor S., entendido desde la perspectiva de la primera persona. Partí del supuesto de que el significado personal de los acontecimientos que regresaban compulsivamente a su mente, en lugar de los recuerdos que buscaba, arrojaría algo de luz sobre el mecanismo de sus fabulaciones y, por tanto, abriría nuevas vías para influir en ellas. Por eso le preguntaba a su esposa, por ejemplo, si el señor S. había pertenecido realmente a equipos de remo y de rugby cuando era estudiante y si de verdad prestaba servicios de asistencia para problemas electrónicos.

Así llegaron a mi conocimiento dos datos que resultaron pertinentes para entender sus fabulaciones. El primero, que había padecido problemas crónicos en los dientes —problemas que finalmente se trataron (con éxito) recurriendo a implantes dentales—; y el segundo, que sufría una arritmia cardíaca controlada mediante un marcapasos.

A continuación reproduzco la transcripción de los primeros minutos de mi décima sesión con el señor S. He elegido este breve fragmento de la grabación porque, cuando aquel día fui a buscarlo a la sala de espera, por un momento pareció reconocer —por primera vez— quién era yo y saber por qué lo atendía. Cuando entré en la sala de espera, se tocó la cicatriz de la craneotomía en la parte superior de la cabeza y dijo: «Hola, doctor». Al entrar con él en la consulta albergaba la esperanza de aprovechar aquel posible destello de lucidez.

Yo: Cuando nos hemos visto en la sala de espera se ha tocado la cabeza.

Señor S.: Creo que el problema es que falta un cartucho. Lo que hay que… con las especificaciones ya estaría. ¿Qué era? ¿UN C49? ¿Lo encargamos?

Yo: ¿Qué hace un cartucho C49?

Señor S.: Memoria. Es un cartucho de memoria; un implante de memoria. Pero nunca llegué a entenderlo. De hecho, hace cinco o seis meses que no lo uso. Parece que realmente no lo necesitamos. Lo cortó todo un médico. ¿Cómo se llamaba? Doctor Solms, creo. Pero parece que realmente no lo necesito. Los implantes funcionan bien.

Yo: Sabe que algo no va bien con su memoria, pero…

Señor S.: Sí, no funciona al cien por cien, pero en realidad no lo necesitamos, solo le faltaban algunos latidos. El análisis mostró que faltaba algo de C o C09. Denise [su primera mujer] me trajo aquí para ver a un médico. ¿Cómo se llamaba? Doctor Solms o algo así. Me hizo uno de esos trasplantes de corazón y ahora funciona bien; late a la perfección…

Yo: Usted sabe que algo va mal. Faltan algunos recuerdos y eso lógicamente le preocupa. Espera que yo pueda arreglarlo, como los otros médicos que le arreglaron los problemas con los dientes y con el corazón. Pero lo desea tanto que le cuesta aceptar que aún no esté arreglado.

Señor S.: Ah, ya veo. Sí, no funciona al cien por cien. [Se toca la cabeza]. Me dieron un golpe en la cabeza. Salí del campo unos minutos pero ahora ya está bien. Supongo que no debería volver. Pero ya me conoce; no me gusta caer. Así que le pregunté a Tim Noakes [un reputado médico deportivo sudafricano] —tengo seguro, ¿sabe?, así que ¿por qué no usarlo?, ¿por qué no acudir al mejor?— y me dijo: «Bien, sigue jugando».

Interrumpiré ahí la escena. Creo que en ella se pueden reconocer fácilmente los trastornos meramente cognitivos de búsqueda y seguimiento de recuerdos que he mencionado antes. Cuando el señor S. me vio entrar en la sala de espera para aquella décima sesión, mi aparición suscitó en él un montón de asociaciones (con médicos, con su cabeza, con la pérdida de memoria, con las intervenciones quirúrgicas, etc.). Sin embargo, en ninguno de los casos recuperó el recuerdo preciso que buscaba. No daba en el blanco, pero casi: recuerdos que pertenecían a las mismas categorías semánticas generales que los recuerdos buscados, pero mal ubicados en el espacio y en el tiempo. La idea de «médico», por ejemplo, suscitaba asociaciones relacionadas con el neurocirujano y con un famoso

médico deportivo en lugar del recuerdo buscado: yo; la idea de «cabeza» evocaba un incidente de conmoción cerebral en lugar de un tumor cerebral; la «pérdida de memoria», un cartucho electrónico en lugar de su amnesia; las «intervenciones quirúrgicas», sus anteriores operaciones dentales y cardiológicas en lugar de la reciente cirugía cerebral, y así sucesivamente. El déficit de seguimiento también queda bien expuesto: el señor S. aceptaba la veracidad de sus recuerdos erróneos con demasiada facilidad. Que se viera a sí mismo como un estudiante de veintitantos años en un campo de rugby (a pesar de todas las pruebas en contra) es un claro ejemplo de ello, al igual que su creencia de que seguía en Johannesburgo.

Ahora bien, si consideramos las fabulaciones del señor S. desde el punto de vista subjetivo, veremos aún más cosas. Imaginen lo que sentirían al darse cuenta de pronto de que no reconocen al médico que acaba de entrar en la habitación, aunque todo indique que es quien los atiende; que no saben en qué habitación están (ni siquiera en qué ciudad); que tienen una enorme cicatriz en la parte superior de la cabeza y no saben por qué; que —de hecho— no recuerdan qué ha pasado hace tan solo dos minutos y mucho menos los días y meses anteriores al momento presente. Probablemente sentirían algo parecido al pánico y se preguntarían si esos fallos de la memoria no se deben a alguna operación en la cabeza que les ha hecho ese médico. Eso es lo que la falta de mecanismos de búsqueda y control de los recuerdos hace sentir al sujeto intencional de la mente, el yo vivo.

Veamos ahora qué hizo el señor S. como consecuencia de los sentimientos que he mencionado (o, dicho de otra forma, qué efectos causales tuvieron en su cognición). Al darse cuenta de que le faltaba su «cartucho de memoria», se tranquiliza (ilusoriamente) diciéndose a sí mismo que «basta con pedir uno nuevo». No convencido del todo, cambia de opinión: en realidad, no necesita el cartucho, se las arregla bien sin él y lleva meses haciéndolo. Entonces establece una asociación entre el cartucho que falta y la cicatriz de la craneotomía; al parecer, un médico le ha cortado algo. Espera que no sea el médico que tiene delante y espera, además, que la operación no haya sido una chapuza. Llegado a ese punto, el señor S. recuerda en paralelo que sus operaciones de odontología y cardiología habían salido bien y confunde (ilusoriamente) aquellas intervenciones con la actual: «salieron bien», «los implantes funcionan bien» y el corazón «late a la perfección». Cuando yo le hago dudar, cambia de táctica. Admite que no funciona al cien por cien, pero al mismo tiempo decide que lo que le ha pasado en la cabeza no ha sido una operación, sino «tan solo una conmoción cerebral»; que está sufriendo los efectos temporales de un accidente deportivo menor y que por eso ha salido del campo unos

minutos. Pero, felizmente, como tiene acceso al mejor médico deportivo que el dinero puede comprar, vuelve a tranquilizarse: puede seguir jugando. Todo irá bien.

En el momento en que nos planteamos las fabulaciones del señor S. en primera persona sale a la luz algo nuevo; el contenido de sus recuerdos erróneos tiene una motivación tendenciosa. Lejos de ser errores de búsqueda aleatorios, contienen un sesgo claro e interesado; tienen el objetivo y el propósito de reconducir su estado de ansiedad a un estado tranquilizador, seguro y conocido, lo cual quiere decir que, tal y como Freud infirió en el caso de los sueños, las fabulaciones están motivadas. Los procesos mentales en la amnesia fabuladora son anhelantes. Sin embargo, esto solo se ve cuando se tiene en cuenta el contexto emocional y el significado personal (experimentado solo por el señor S.) de los implantes dentales («los implantes funcionan bien») y de los marcapasos cardíacos («late a la perfección»), como haría un psicoanalista. Eso es lo que los neuropsicólogos no ven cuando pretenden ser totalmente objetivos; o como dijo Sacks, cuando excluyen la psique.

La observación que acabo de describir adoptando la perspectiva de la primera persona también revela algo nuevo sobre el mecanismo de la fabulación, algo que se pasa por alto cuando se observa desde el punto de vista de la tercera persona. En efecto, nos dice que la fabulación es consecuencia no solo de déficits en la búsqueda estratégica y en el seguimiento de la fuente (es decir, los «cartuchos de memoria» que faltan), sino también de una desinhibición de formas de recuerdo con mayor mediación emocional, como ocurriría en la memoria de un niño. Este mecanismo psicodinámico tiene implicaciones para el tratamiento de la fabulación y, por supuesto, de cara a saber qué procesos cerebrales están implicados. El caso es que las funciones de búsqueda y seguimiento precisas de la memoria dependen en parte de los circuitos colinérgicos del prosencéfalo basal, que limitan los mecanismos de «recompensa» del circuito dopaminérgico mesocortical-mesolímbico en la recuperación de recuerdos. De hecho, en los sueños se produce una liberación similar de la búsqueda dopaminérgica.[62] Por eso, cuando informé a mis colegas del caso del señor S., lo titulé «El hombre que vivía en un sueño».

Esto me permitió, al igual que en el caso de los sueños, vincular provisionalmente el mecanismo dopaminérgico no restringido de «recompensa», «querer» o «BÚSQUEDA» con la noción de Freud de «realización de deseos»,[63] un concepto metapsicológico estrechamente relacionado con su concepto de «pulsión».[64] A la inversa, las funciones de los núcleos colinérgicos del prosencéfalo

pueden asociarse en algunos aspectos con las influencias inhibitorias de la «prueba de realidad».[65] De esta forma, empecé a traducir las inferencias de Freud sobre los mecanismos funcionales de la subjetividad a sus equivalentes fisiológicos.

Esos fueron mis primeros pasos, pero es evidente que no podemos basar semejantes generalizaciones en meras pruebas clínicas de un solo caso. Tras formular mi impresión sobre el señor S., recurrí a evaluadores «ciegos» (colegas que desconocían mi hipótesis) para medir, en una escala de Likert de siete puntos, el grado de agradabilidad frente a desagradabilidad en una muestra continua no seleccionada de 155 de sus fabulaciones. Los resultados fueron estadísticamente (muy) significativos: al compararlas con los recuerdos diana a los que sustituían, las fabulaciones del señor S. mejoraban considerablemente su situación desde el punto de vista emocional.[66] Más adelante, junto a mis colaboradores de investigación, demostramos que se producía un efecto igual de fuerte en estudios con muchos otros pacientes que sufrían fabulaciones. En estudios empíricos posteriores, los efectos reguladores del estado de ánimo de la fabulación que inferí en el caso clínico del señor S. quedaron estadísticamente validados.[67] Este programa de investigación abrió un planteamiento completamente nuevo de la neuropsicología de la fabulación[68] y de trastornos relacionados como la anosognosia.[69] También sentó las bases para un enfoque novedoso de trastornos psiquiátricos comunes, como la adicción y la depresión mayor.[70] He pasado treinta años desarrollando este enfoque «neuropsicoanalítico» de la enfermedad mental e intentando devolver la subjetividad a la neurociencia.[71]

Mientras seguía mi formación psicoanalítica y acumulaba experiencias clínicas similares a la que he descrito, me invitaron a exponer mis hallazgos en una serie de presentaciones científicas en Nueva York. La primera fue un simposio de un solo día celebrado en 1992 en la Academia de Medicina de Nueva York, al que siguieron seminarios mensuales con mis colaboradores más cercanos en la Sociedad e Instituto Psicoanalíticos de Nueva York.[72] Las reuniones de este grupo de colegas fueron dando lugar a actividades parecidas en muchos otros rincones del mundo, lo que nos llevó a la decisión, en 1999, de crear una nueva revista que nos sirviera de vehículo de comunicación. Como la revista necesitaba un nombre, pude estrenar mi término inventado, Neuropsychoanalysis.

Mi trabajo en este campo interdisciplinar fue ampliamente respaldado

por Eric Kandel, a quien conocí en 1993. Kandel se distingue de la mayoría de sus colegas neurocientíficos por el gran aprecio que siente por Freud. De hecho, su intención inicial era formarse como psicoanalista, pero Ernst Kris, uno de los analistas más importantes de la época y padre de la que entonces era su novia, le disuadió por no tener —según ha contado el propio Kandel— una personalidad adecuada para la práctica clínica de la psiquiatría. Yo diría que Kandel agradece que el consejo del anciano Kris lo orientara hacia la investigación del cerebro.

Cinco años después de conocernos (y dos años antes de ganar el Premio Nobel), Kandel publicó un artículo titulado «Un nuevo marco intelectual para la psiquiatría», en el que defendía que la psiquiatría del siglo XXI debería basarse en la integración de la neurociencia y el psicoanálisis.[73] En un segundo artículo afirmó: «El psicoanálisis sigue representando la visión más coherente y más satisfactoria, intelectualmente hablando, de la mente».[74] Ahí coincidimos por completo: con todos sus defectos, el psicoanálisis nos ofrece en estos momentos el mejor punto de partida conceptual para un abordaje científico de la subjetividad.

No es de extrañar, por tanto, que Kandel aceptara mi invitación a unirse al consejo editorial fundador de Neuropsychoanalysis, junto con una masa crítica de otros destacados neurocientíficos y psicoanalistas que también consideraron que nuestras disciplinas debían avanzar por ese camino.[75] Un año más tarde, al despuntar el nuevo siglo, fundamos la Sociedad Internacional de Neuropsicoanálisis, con Jaak Panksepp y conmigo como sus primeros copresidentes. Lo hicimos durante el congreso inaugural de la sociedad, que desde entonces se reúne cada año en distintas ciudades del mundo. El tema del primer congreso fue la emoción. La reunión se celebró en el Royal College of Surgeons of England y los ponentes plenarios fueron Oliver Sacks, Jaak Panksepp, Antonio Damasio y yo mismo.

Ya he mencionado antes mi relación con Oliver Sacks y también el libro Neurociencia afectiva de Jaak Panksepp, título que aludía a la poca atención que la neurociencia cognitiva prestaba al «afecto» (affect, el término técnico en inglés para los sentimientos). Tras leer su libro, inicié con Panksepp una estrecha colaboración científica que durante las dos décadas siguientes me hizo desplazar de forma definitiva el foco de mi trabajo desde la corteza al tronco del encéfalo. Estoy profundamente en deuda con él por haberme mostrado el camino hacia las ideas que expondré en las páginas siguientes, y por eso este libro está dedicado a su memoria.

La primera vez que entré en contacto con la obra de Antonio Damasio y de su esposa, Hanna Damasio, fue durante mi formación neuropsicológica. Ambos eran neurocientíficos cognitivos muy respetados y su libro de texto Lesion Analysis in Neuropsychology (1989) fue una ayuda indispensable en mi investigación sobre los sueños. Sin embargo, el libro que dio fama mundial a Damasio fue El error de Descartes (1994), un apasionado alegato a favor de un mayor reconocimiento del afecto en la neurociencia cognitiva.

Panksepp y Damasio también tuvieron un papel importante en el XXII Congreso Internacional de Neuropsicoanálisis de 2011, que supuso un punto de inflexión para esta disciplina. Celebrado en Berlín, el tema del congreso fue el cerebro corporizado. Los otros ponentes plenarios fueron Bud Craig y Vittorio Gallese, dos de los mayores expertos mundiales en el tema. Mi papel en estos congresos suele consistir en una alocución final para resumir los principales temas tratados y, sobre todo, integrar las perspectivas neurocientíficas y psicoanalíticas presentadas. En esa ocasión, mi tarea resultó especialmente ardua.

El primer obstáculo fue el agudo enfrentamiento entre Damasio y Craig durante el congreso respecto a cómo se genera en el cerebro un «yo» sintiente. Aunque ambos científicos coincidían en que el sentido del yo surge de regiones cerebrales que supervisan los estados corporales, Damasio —siguiendo a Panksepp— defendía que los mecanismos en cuestión se hallaban, al menos en parte, en el tronco encefálico. Craig, en cambio, afirmaba que estaban solo en la corteza, en concreto en la ínsula anterior. Este desacuerdo fue relativamente fácil de resolver en mi discurso de clausura, porque Damasio había proporcionado datos convincentes, centrándose en un paciente con extirpación completa de la ínsula cortical. Describiré a ese paciente en el próximo capítulo.

Fue mucho más difícil de conciliar una contradicción de peso que surgió durante el congreso entre los nuevos planteamientos de Panksepp y Damasio, por un lado, y los antiguos puntos de vista de Freud, por el otro.

El paciente de Damasio que carecía de corteza insular «informaba de sensaciones de hambre, sed y deseo de evacuar, y se comportaba en consecuencia».[76] Estas sensaciones son ejemplos de lo que Panksepp denomina «afectos homeostáticos», afectos que regulan las necesidades vitales del cuerpo. Freud los denominó «pulsiones», la fuente de su «energía psíquica», el «impulso primario del mecanismo psíquico». El

término amplio que empleó Freud para denominar la parte de la mente que realiza estas funciones vitales es id (ello):

El ello, aislado del mundo exterior, tiene su propio mundo de percepción. Detecta con extraordinaria agudeza ciertas alteraciones en su interior —en particular, las oscilaciones en la tensión de sus necesidades pulsionales—, y estas alteraciones devienen conscientes como sensaciones de la serie placer-displacer. A ciencia cierta, es difícil decir por qué vías y con ayuda de qué órganos terminales sensibles se producen estas percepciones. Pero es un hecho comprobado que las autopercepciones —las sensaciones cenestésicas y las sensaciones de placer-displacer— gobiernan el paso de los acontecimientos en el ello. El ello obedece al inexorable principio del placer.[77]

Como bien recordarán, mi objetivo científico era traducir esos conceptos metapsicológicos a los lenguajes de la anatomía y la fisiología para así poder integrar el planteamiento de Freud en la neurociencia. Pero aquí había tropezado con una contradicción de peso en la concepción clásica de Freud, quien había llegado a la conclusión de que el «ello» era inconsciente, una de sus concepciones más fundamentales sobre el funcionamiento de la mente. Para mí estaba claro que la parte del encéfalo que mide la «demanda de trabajo que se hace a la mente a consecuencia de su conexión con el cuerpo» —la parte que genera lo que Freud llamaba «pulsiones», sinónimo de los «afectos homeostáticos» de Panksepp (que desencadenan su mecanismo de BÚSQUEDA anhelante)— se halla en el tronco encefálico y en el hipotálamo (véase fig. 1). Esa es la parte del encéfalo que obedece al «principio de placer». Pero ¿cómo pueden ser inconscientes los sentimientos de placer? Como hemos visto con el paciente de Damasio, pulsiones como el hambre y la sed y el deseo de evacuar se sienten. Por supuesto que se sienten. Sin embargo, Freud dijo que el ello —la sede de las pulsiones— era inconsciente. Freud se había nutrido de la misma doctrina clásica que Craig (como yo mismo, al menos al principio) y, por tanto, había situado la conciencia en la corteza cerebral. Así, en el ensayo de 1920 ya citado, cuando esperaba que las deficiencias de sus teorías pudieran solucionarse en cuanto estuviéramos en condiciones de reemplazar los términos psicológicos por otros fisiológicos y químicos, escribió lo siguiente:

Lo que la conciencia produce consiste esencialmente en percepciones de excitaciones procedentes del mundo exterior y de sentimientos de placer y displacer que solo pueden surgir del aparato mental; por ello es posible atribuir al sistema P-Cc [conciencia perceptual] una posición en el espacio. Tiene que hallarse en la frontera entre lo interior y lo exterior, estar vuelto hacia el mundo exterior y envolver a los otros sistemas psíquicos. Veremos que estas suposiciones no tienen nada atrevidamente nuevo; nos hemos limitado a adoptar las ideas de localización de la anatomía cerebral, que sitúa la «sede» de la conciencia en la corteza cerebral, en el estrato más exterior, envolvente, del órgano central. La anatomía cerebral no necesita ocuparse de la razón por la cual —dicho en términos anatómicos— la conciencia está ubicada justamente en la superficie del encéfalo, en vez de estar alojada en alguna otra parte, en lo más recóndito de él. [78]

Añadiré otra cita de Freud por si queda alguna duda sobre su convicción de que toda la conciencia (incluidas las sensaciones de placer y displacer) se halla en la corteza:

El proceso de algo que deviene consciente está vinculado, sobre todo, a las percepciones que nuestros órganos sensoriales reciben del mundo exterior. Desde el punto de vista topográfico, pues, es un fenómeno que sucede en el estrato cortical más exterior del yo. Es cierto que también recibimos noticias conscientes del interior del cuerpo, los sentimientos, que de hecho ejercen una influencia más perentoria sobre nuestra vida mental que las percepciones externas; asimismo, bajo ciertas circunstancias, también los órganos de los sentidos brindan sentimientos, sensaciones de dolor, además de sus percepciones específicas. Sin embargo, dado que estas sensaciones —como se las llama para distinguirlas de las percepciones conscientes— parten también de los órganos terminales, y a todos estos los concebimos como prolongación, como unos emisarios del estrato cortical, podemos mantener la afirmación anterior. La única distinción sería que, en el caso de los órganos terminales de las sensaciones y los sentimientos, el propio cuerpo sustituiría al mundo exterior.[79]

Queda claro que para Freud los sentimientos conscientes, no menos que las percepciones, se generan en el «yo» (la parte de la mente que

él identificaba con la corteza),[80] no en el «ello» inconsciente, que ahora me veía obligado a ubicar en el tronco encefálico y el hipotálamo. Parecía, pues, que Freud había interpretado al revés la relación funcional entre el «ello» (tronco encefálico) y el «yo» (corteza), al menos en lo que respecta a los sentimientos. Él creía que el yo perceptor era consciente y el ello sintiente era inconsciente. ¿No tendría al revés su modelo de la mente?[81]

[36] Ese año me aceptaron como estudiante, pero no comencé hasta 1989.

[37] De hecho, Freud no le puso ningún título a aquel manuscrito inédito; el título se lo inventaron los traductores ingleses. En su correspondencia con Wilhelm Fliess, Freud lo llamó «Psicología para neurólogos», «Esbozo de una psicología» y «la Psicología».

[38] Carta a Hallmann, 1842, publicada en Du Bois-Reymond, 1918, p. 108. También se cita a menudo el prólogo de Du Bois-Reymond a su Über die Lebenskraft (1848-1884, pp. xliii-xliv): «En los organismos y sus partículas no hay ninguna fuerza nueva funcionando, ninguna que no esté también en funcionamiento fuera de ellos. Tampoco hay fuerzas que merezcan el nombre de "fuerza vital". La separación entre las llamadas naturalezas orgánica e inorgánica es completamente arbitraria».

[39] Bechtel y Richardson, 1998.

[40] Véase mi análisis de este término en Solms (en prensa).

[41] Freud, 1950b, p. 295.

[42] La prioridad de Freud por formular la posición «funcionalista» no goza de reconocimiento general (Freud, 1900, p. 536): «[…] en nuestro intento de hacer inteligibles las complicaciones del funcionamiento mental, diseccionando la función y asignando sus diferentes constituyentes a las distintas partes del aparato. Que yo sepa, hasta ahora no se ha hecho el experimento de emplear este método de disección para investigar de qué forma se compone el instrumento mental, y no veo nada malo en ello». Cf. Shallice, 1988.

[43] Freud utilizó por primera vez este curioso término para, según él, referirse a un nivel de explicación que incorpora tanto la psicología como la biología (carta a Fliess del 10 de marzo de 1898; Freud, 1950a). Como Freud escribió una vez a Georg Groddeck: «El inconsciente es el eslabón perdido entre lo físico y lo mental» (carta del 5 de junio de 1917). Véase Solms, 2000b. Véase también mi presentación en un encuentro de la Academia de Ciencias de Nueva York celebrado para conmemorar el centenario del «Proyecto» de Freud (Solms, 1998). En aquel congreso habló el gran Karl Pribram y pude conocer también al pionero de la neurofisiología Joseph Bogen. Recuerdo muy bien que, como quien no quiere la cosa, dijo que la conciencia la generaban los núcleos intralaminares del tálamo; fue la primera vez que oí a alguien sugerir que la corteza no es intrínsecamente consciente. Véase Bogen, 1995.

[44] La crítica de Freud (1891) al localizacionismo sentó las bases del planteamiento de los «sistemas funcionales» que imperó después en la neuropsicología y, más tarde, en el cognitivismo. Trato esta cuestión en detalle en Solms y Saling, 1986; y en Solms, 2000b.

[45] De ahí el título del libro de Frank J. Sulloway (1979), Freud: Biologist of the Mind.

[46] Freud, 1914, p. 78.

[47] Freud, 1920, p. 83.

[48] Carta a Fliess de 20 de octubre de 1895.

[49] Carta a Fliess de 29 de noviembre de 1895.

[50] Carta a Fliess de 25 de mayo de 1895.

[51] Actualmente estoy preparando una traducción al inglés de The Complete Neuroscientific Works of Sigmund Freud (en 4 volúmenes). Véase también mi revisión de las traducciones (al inglés) de Strachey: The Revised Standard Edition of the Complete Psychological Works of Sigmund Freud (24 volúmenes).

[52] Freud, 1950b, pp. 303, 316. Muchos años antes que Freud, Baruch Spinoza escribió que el deseo es la esencia misma del ser humano.

[53] Freud, 1901, p. 259.

[54] Freud, 1920, p. 60.

[55] Freud, 1915a, pp. 121-122; la cursiva es mía.

[56] Carta a Fliess de 25 de mayo de 1895.

[57] Freud, 1940, p. 197.

[58] Esta figura es una adaptación de Braun et al., 1997. El estudio de Braun era puramente descriptivo y sus hallazgos son compatibles con la teoría de Freud, pero no la confirman de forma experimental porque no probaron ninguna de las predicciones que se derivaban de ella. Sí lo hizo en cambio una alumna mía (Catherine Cameron-Dow, 2012), que sometió a prueba hace poco la teoría de Freud de que los sueños protegen el sueño. Su confirmación de la hipótesis es objeto de un estudio más amplio de mi colega de Berlín Tamara Fischmann (en curso).

[59] Por cierto, el debate lo presidió nada menos que David Chalmers. El resultado revirtió una votación producida en 1978 después de que Hobson presentara su teoría de «activación-síntesis» a los miembros de la Asociación Americana de Psiquiatría allí congregados.

[60] Solms y Saling, 1986.

[61] Véase Braun, 1999.

[62] Malcolm-Smith et al., 2012.

[63] Solms, 2000c; Solms y Zellner, 2012.

[64] Cabe observar que, en la época en la que desarrolló su concepto de «pulsión libidinal», Freud era un consumidor habitual de cocaína, un alcaloide que activa mucho el sistema dopaminérgico de BÚSQUEDA. Eso daría lugar a la hipótesis —fantasiosa o no tanto— de que haber experimentado en persona los efectos motivacionales generalizados de la cocaína podría haber contribuido a su reconocimiento de la existencia de ese mecanismo motivacional polivalente en la mente.

[65] En el capítulo 7 relacionaré la «realización de deseos» con la codificación predictiva, y la «prueba de realidad» con lo que hoy se denomina «error de predicción» (o error de predicción de precisión modulada).

[66] Fotopoulou, Solms y Turnbull, 2004.

[67] Turnbull, Jenkins y Rowley, 2004.

[68] Fotopoulou y Conway, 2004; Turnbull, Berry y Evans, 2004; Fotopoulou et al., 2007, 2008a,b; Turnbull y Solms, 2007; Fotopoulou, Conway y Solms, 2007; Fotopoulou, 2008, 2009, 2010a,b; Coltheart y Turner, 2009; Cole et al., 2014; Besharati, Fotopoulou y Kopelman, 2014; Kopelman, Bajo y Fotopoulou, 2015.

[69] Véase la revisión en Turnbull, Fotopoulou y Solms, 2014. Véase también Besharati et al., 2014, 2016.

[70] Zellner et al., 2011.

[71] Solms y Turnbull, 2002, 2011; Panksepp y Solms, 2012; Solms, 2015a.

[72] Al final, aquellas presentaciones se recopilaron en forma de volumen: Kaplan-Solms y Solms, 2000.

[73] Kandel, 1998.

[74] Kandel, 1999, p. 505.

[75] Estos neurocientíficos eran Allen Braun, Jason Brown, Antonio Damasio, Vittorio Gallese, Nicholas Humphrey, Eric Kandel, Marcel Kinsbourne, Joseph LeDoux, Rodolfo Llinás, Georg Northoff, Jaak Panksepp, Michael Posner, Vilanayur Ramachandran, Oliver Sacks, Todd Sacktor, Daniel Schacter, Carlo Semenza, Tim Shallice, Wolf Singer y Max Velmans. Entre los psicoanalistas se encontraban Peter Fonagy, Andre Green, Ilse Grubrich-Simitis, Otto Kernberg, Marianne Leuzinger-Bohleber, Arnold Modell, Barry Opatow, Allan Schore, Theodore Shapiro, Riccardo Steiner y Daniel Widlöcher.

[76] Damasio, Damasio y Tranel, 2013.

[77] Freud, 1940, p. 198. En todo el libro utilizo mis versiones revisadas de las traducciones al inglés de James Strachey (véase Solms, en prensa).

[78] Freud, 1920, p. 24; la cursiva es mía.

[79] Freud, 1940, pp. 161-162; la cursiva es mía.

[80] Véase Freud, 1923, p. 26: «El yo es, ante todo, un ser corpóreo, y no una mera entidad superficial, sino incluso la proyección de una superficie. Si queremos encontrarle una analogía anatómica, lo mejor será identificarlo con el "homúnculo cortical" de los anatomistas, que se halla cabeza abajo sobre la corteza cerebral, tiene los pies hacia

arriba, mira hacia atrás y ostenta, como sabemos, la zona de la palabra en el lado izquierdo, […]. El yo se deriva en última instancia de las sensaciones corporales, principalmente de aquellas que brotan en la superficie del cuerpo, por lo que puede considerarse al yo como una proyección mental de dicha superficie».

[81] Véase Solms, 2013.

03

La falacia cortical

A finales de 2004, el neurocientífico Björn Merker se reunió con cinco familias de niños con discapacidades neurológicas para pasar una semana juntos en Disney World. La edad de los niños iba de los diez meses a los cinco años. Los visitantes montaron en varias atracciones y la preferida fue el paseo en barca que lleva por nombre It's a Small World After All. Todos se fotografiaron junto a Mickey Mouse, comieron palomitas, mazorcas de maíz y helado —puede que un poco más de lo necesario— y bebieron un montón de refrescos. En algún momento dieron muestras de estar desbordados; en otros hubo lágrimas. Sin embargo, pese a esos ratos de tensión, lo que impresionó al doctor Merker fue ver lo bien que se lo pasaban y lo contentos que estaban de estar allí. Porque, según una de las premisas más básicas de la neurología, esos niños deberían haberse hallado en «estado vegetativo», a un paso del coma, presentando como únicas funciones autónomas la regulación de los latidos del corazón, la respiración y la actividad gastrointestinal, y como únicas respuestas motoras reflejos simples como parpadear y tragar.

Merker había conocido a esas familias un año antes, cuando se apuntó a un grupo de autoayuda para cuidadores de niños afectados por una rara enfermedad cerebral, la hidranencefalia. Desde principios de 2003, Merker leyó más de veintiséis mil mensajes de correo electrónico cruzados entre los miembros del grupo. Le preocupaba que la suposición general —si nos basamos en la teoría— de que esos niños serían «vegetativos» pudiera haberse convertido en una profecía autocumplida, causada no por la enfermedad en sí, sino por el hecho de que la mayoría de los pediatras y neurólogos los trataban como si fueran completamente insensibles. He aquí lo que informó:

Estos niños no solo están despiertos y a menudo alerta, sino que también presentan capacidad de respuesta a su entorno en forma de reacciones emocionales o de orientación a lo que ocurre a su alrededor. […] Además, tienen crisis de epilepsia de ausencia. Los padres distinguen en sus hijos estas interrupciones de accesibilidad y las comentan con expresiones como «se ha ido un momento a hablar

con los ángeles», y reconocen enseguida cuando su hijo o hija ha regresado. […] El hecho de que estos niños presenten esa clase de episodios sería una prueba de peso de su estado consciente.[82]

La observación más importante de Merker no es que los pacientes pierdan y recuperen el estado de alerta, sino que muestren «capacidad de respuesta a su entorno en forma de reacciones emocionales o de orientación a lo que ocurre a su alrededor». Es decir, el rasgo definitorio de algo de lo que se supone que carecen los pacientes vegetativos: intencionalidad. Por eso el estado vegetativo también se define como «vigilia sin respuesta».[83] Frente a la creencia generalizada de que estos pacientes registran los estímulos visuales, auditivos y táctiles de forma inconsciente, Merker vio a niños que expresaban placer mediante risas y sonrisas, así como aversión mediante pataletas, espaldas arqueadas y llantos: «sus rostros se veían animados por esos estados emocionales». Observó también que los adultos aprovechaban la capacidad de respuesta de sus niños para crear secuencias de juego en las que se pasaba de forma predecible de la sonrisa a las risas, y de estas a las carcajadas y a una gran excitación de los niños. Sus respuestas eran más intensas ante las voces y acciones de sus padres y de otras personas conocidas, y mostraban preferencia por determinadas situaciones frente a otras. Por ejemplo, parecían disfrutar con juguetes, melodías o vídeos concretos, e incluso llegaban a esperar que se incluyeran en la rutina diaria. Aunque no todos los comportamientos eran iguales, algunos niños mostraban claramente iniciativa (hasta donde les permitían sus discapacidades motoras); por ejemplo, dando patadas a sonajeros que colgaban de un marco especial construido a tal efecto o activando mediante interruptores algún juguete favorito. Dichos comportamientos iban acompañados de señales de placer o agitación situacionalmente apropiadas.

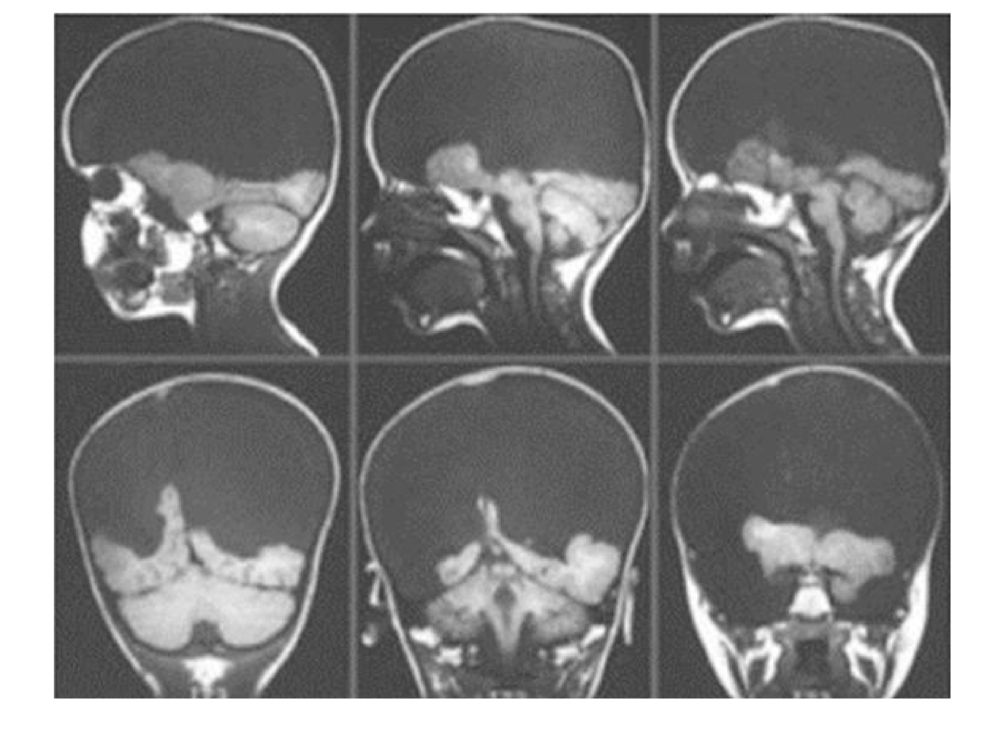


Figura 4

. Resonancia magnética del encéfalo de una niña de tres años que nació sin corteza cerebral. La gran región oscura dentro del cráneo indica la ausencia de tejido.

Queda claro que no se puede describir a estos niños como «vegetativos». Lo que hace que estos casos resulten tan sorprendentes es que los niños hidranencefálicos nacen sin corteza cerebral, normalmente a consecuencia de un derrame cerebral masivo intrauterino que provoca la reabsorción del prosencéfalo, tras lo cual el cráneo del bebé se inunda de líquido cefalorraquídeo en lugar de tejido cerebral (de ahí el término hidranencefalia, que significa «agua en lugar de encéfalo»). Podemos ver una ilustración en la figura 4, una resonancia magnética del encéfalo de una niña de tres años que nació sin corteza cerebral. En la figura 5 (p. 70) se ve la respuesta emocional de esa misma niña cuando le ponen a su hermano bebé en el regazo.

¿En qué parte del encéfalo se genera la conciencia? Durante los últimos ciento cincuenta años, la respuesta casi universal a esta pregunta era «en la corteza». Era el único punto en el que Freud y la tradición imperante en la ciencia mental del siglo XX estaban de

acuerdo. De ser así, sin embargo, en ausencia de la corteza, la conciencia debería desaparecer, pero no parece ser el caso de los niños hidranencefálicos. De hecho, todas las pruebas conductuales sugieren que son conscientes. No están en coma ni tampoco viven en estado vegetativo.



Figura 5

. Reacción de una niña hidranencefálica cuando le ponen a su hermano bebé en el regazo.

¿Hay que rechazar entonces la teoría cortical de la conciencia? No nos precipitemos. Se podría objetar, por ejemplo, que en esos niños no hubo extirpación quirúrgica de la corteza, el procedimiento llamado «decorticación».

Veamos qué ocurre tras una decorticación. Como es obvio, estos procedimientos experimentales no pueden hacerse con bebés humanos, pero sí que se han realizado con otros mamíferos recién nacidos, como perros, gatos y ratas. Y siempre con los mismos resultados: según los criterios conductuales objetivos que normalmente utilizamos para medirla, la conciencia se conserva. (Por irónico que parezca, el resultado de experimentos como estos fue uno de los motivos que nos llevaron a un cambio de actitud respecto de la ética de este tipo de investigaciones). El comportamiento posoperatorio de estos animales no se ajusta ni siquiera un poco a la definición de «comatoso» o «vegetativo». Merker escribe que no presentan «ninguna anomalía grave en el comportamiento que permita

a un observador fortuito identificarlos como discapacitados». Antonio Damasio coincide: «Los mamíferos decorticados muestran una notable persistencia de comportamiento coherente y orientado a objetivos que es compatible con los sentimientos y la conciencia».[84] Las ratas decorticadas al nacer, por ejemplo, se aguantan en pie a cuatro patas, se incorporan sobre las patas traseras, trepan, se cuelgan de barras y adoptan posturas normales para dormir. Se acicalan, juegan, nadan, comen y se defienden. Ambos sexos son capaces de aparearse con éxito cuando se emparejan con compañeros de jaula normales. Cuando crecen, las hembras muestran lo esencial del comportamiento materno, que, aunque deficiente en algunos aspectos, les permite criar la camada hasta la madurez.[85]

La situación es aún más extraña de lo que parece a primera vista. En muchos aspectos, los mamíferos decorticados son, de hecho, más activos, emocionales y reactivos que los normales. Panksepp solía presentar a sus estudiantes de posgrado dos grupos de ratas para que señalaran, basándose en su comportamiento, cuáles habían sido operadas. Por lo general, los estudiantes señalaban las normales, aduciendo que el otro grupo (las decorticadas) era «más vivo».[86]

Si seguimos la hipótesis de que la conciencia reside en la corteza, entonces esos animales tan vivaces —y los niños expresivos y con respuestas emocionales que Merker observó en Disney World— deberían ser, en cierto sentido, inconscientes. ¿Cómo se explica eso?

Existe una respuesta convencional para esta pregunta: hay, por decirlo de alguna manera, «dos cerebros», que duplican ciertas funciones y se comunican entre sí en puntos determinados, pero que por lo demás no se parecen en nada. Uno de ellos (la corteza) es psicológico y consciente, mientras que el otro (el tronco encefálico) no es ninguna de las dos cosas. La información procedente de los órganos de los sentidos no llega solo a la corteza, sino también a los tubérculos cuadrigéminos superiores del tronco encefálico a través de un conjunto de conexiones subcorticales (véase fig. 6, p. 73). Estas conexiones procesan la información sensorial, pero lo hacen de forma inconsciente. Un ejemplo sería el conocido fenómeno de la «visión ciega», que se produce cuando se destruye la corteza visual.[87] Estos pacientes son capaces de responder a estímulos visuales, pero cuando se les pide que describan cómo es su «visión», aseguran no experimentar ninguna imagen visual y dicen que adivinan por instinto o corazonada dónde están los estímulos visuales que les presentan (y lo hacen con notable precisión). Veamos el caso de un paciente al que llamaremos «T. N.», relatado por el neurocientífico Lawrence Weiskrantz. Aun siendo completamente ciego —o, dicho de otro

modo, pese a carecer por completo de experiencia visual consciente—, T. N. sorteó con destreza los obstáculos que le habían interpuesto a lo largo de un pasillo. Cuando le preguntaron después, dijo que ni siquiera sabía que estaba evitando algo.[88] Esta sorprendente capacidad es posible porque la vía que va del nervio óptico a los tubérculos cuadrigéminos superiores del tronco encefálico permanece intacta a pesar de la ausencia de corteza occipital.[89]

La existencia de la visión ciega, precisamente, se ha utilizado para deducir que la conciencia de la percepción visual tiene que producirse en la corteza y no en el tronco encefálico, donde se da por hecho que «no hay nadie ahí» porque es una máquina autónoma que procesa la información visual de la misma forma no consciente que lo hace una cámara. Este principio es también aplicable a los demás sentidos, que tienen su respectiva zona de conciencia en la corteza, pero que también transmiten —a excepción del sentido del olfato— información a los tubérculos cuadrigéminos superiores inconscientes del tronco encefálico.

La niña de la figura 5 no tiene corteza funcional. Si hacemos caso al argumento anterior, habrá que suponer que percibe inconscientemente que le colocan a su hermano en el regazo, sin generar ninguna percepción consciente de la situación; incluso que es incapaz de experiencia consciente alguna. Vendría a ser lo que el filósofo David Chalmers denomina una «zombi». Aunque en ciertos aspectos actúa con normalidad, dentro de ella está todo oscuro. (Hay que señalar aquí que el concepto filosófico de «zombi» no coincide con el de Hollywood: se trata de una criatura humanoide imaginaria que actúa en todos los sentidos como si fuera consciente —lo que impide diferenciarla desde fuera de una persona normal—, pero que en realidad carece de la dimensión interior de la experiencia subjetiva).

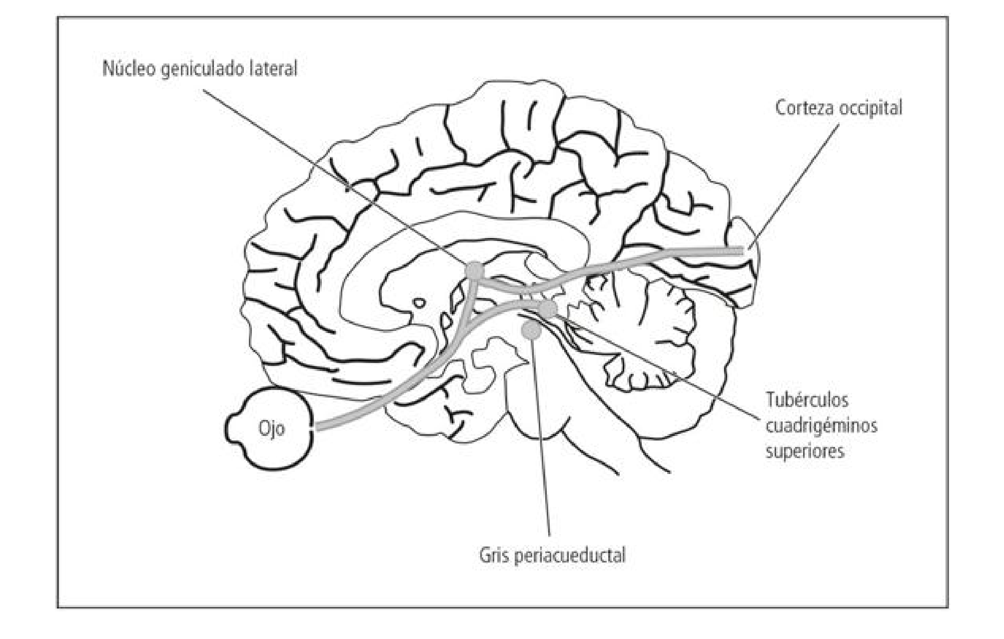


Figura 6

. Esta figura muestra la ubicación de los tubérculos cuadrigéminos superiores y la corteza occipital y sus conexiones con el ojo. Existen conexiones similares para las demás modalidades sensoriales. También se indica la sustancia gris periacueductal (SGP), de la que aún no he hablado, pero cuya importancia veremos enseguida.

En un exhaustivo repaso de las variedades de lo que llamamos «conciencia», el neurólogo Adam Zeman atribuyó dos significados principales al término: «conciencia como estado de vigilia» y «conciencia como experiencia».[90] Desarrollando la idea, Anton Coenen escribió más tarde: «La conciencia en el primer sentido (conciencia como estado de vigilia) es una condición necesaria para la conciencia en el segundo sentido (conciencia como experiencia o conciencia fenoménica)».[91]

Ambos significados coinciden con la distinción convencional en neurología entre el «nivel» cuantitativo y el «contenido» cualitativo de la conciencia. Sería posible entonces que aunque los animales decorticados y los niños hidranencefálicos estén despiertos, su experiencia carezca de contenido. Hemos visto que son reactivos y presentan iniciativa conductual. Sin embargo, podemos mantener la hipótesis de que la sede de la «conciencia como experiencia» es la

corteza si postulamos que ser consciente en el sentido conductual de estar despierto y reactivo es significativamente distinto a tener conciencia en el sentido fenomenológico, es decir, ser un sujeto de experiencia.

Llegados a este punto, si ustedes se parecen mínimamente a mí, deben de sentir bastante incomodidad. La niña de la figura 5 es consciente en el sentido de que está despierta, reacciona e inicia sus propios movimientos dirigidos a un objetivo. Pero si «conciencia» significa tener experiencias fenoménicas, se supone que no tiene nada de eso. Recurriendo a otra expresión popular entre los filósofos, no hay «algo que es como» ser ella.

Ahora me atreveré a explicar qué es lo que me resulta tan incómodo de esta línea de razonamiento, aunque suene un poco ingenuo. Si me baso en mi propia experiencia, estar despierto y reactivo es más o menos lo mismo que tener experiencia consciente. Que yo sepa, nunca estoy despierto y reactivo pero fenoménicamente inconsciente. Ambas cosas van juntas. En cuanto me despierto, me doy cuenta de las cosas. De hecho, desde donde estoy sentado tengo la sensación de que mi conciencia interior provoca mi capacidad de respuesta exterior, al menos hasta cierto punto. En circunstancias normales, cuando me doy cuenta de las cosas —o sea, cuando soy consciente de ellas—, respondo intencionadamente. Es de suponer que a ustedes les ocurre lo mismo.

Esa es la razón por la que medimos la conciencia de los pacientes neurológicos en función de su capacidad de reacción. ¿Qué otra cosa podríamos hacer? En la práctica clínica, distinguimos entre estados como el coma, el estado vegetativo y la vigilia con plena capacidad de reacción utilizando la escala de coma de Glasgow, de quince puntos. Esta escala se compone de pruebas de las respuestas oculares del paciente, sus respuestas verbales a preguntas y sus respuestas motoras a instrucciones y (si es necesario) al dolor.[92] Si el paciente responde plenamente, lo consideramos consciente y lo tratamos en consecuencia. No nos preocupa que puedan responder como si estuvieran conscientes cuando en realidad son zombis. Lo que preocupa es lo contrario: los neurólogos deben tener en cuenta la posibilidad de que los pacientes no respondan de forma externa pero estén conscientes por dentro, como ocurre, por ejemplo, en casos de discapacidad motora completa como el «síndrome de enclaustramiento».

¿Qué hacemos entonces? Es innegable que el problema filosófico de las otras mentes nos hace cuestionar cuánto podemos deducir sobre

los estados subjetivos de los animales y de otras personas a partir de su comportamiento, del mismo modo que el hecho de que alguien mienta o finja puede llevarnos a error sobre sus estados interiores. Lo que no hace el problema de las otras mentes es afirmar que la experiencia interior y la respuesta exterior son independientes entre sí. Dado que nosotros sentimos la presencia de un fuerte vínculo en nuestros propios casos individuales, está claro que la carga de la prueba debe recaer en quienes reivindican que el comportamiento emocionalmente receptivo no implica experiencia fenoménica. Eso es lo que suelen exigir las reglas de la ciencia, y más cuando se trata de seres humanos de los que se puede suponer que —con independencia de otros trastornos neurológicos que puedan sufrir— generan su comportamiento de vigilia de una forma que podríamos llamar «habitual», es decir, mediante el funcionamiento normal de las partes no lesionadas del encéfalo, en lugar de mentir, fingir o recurrir a las artimañas que cabe esperar en robots filosóficos imaginarios.

Todo esto tiene unas implicaciones considerables para la ética médica. No hace mucho, una colega psiquiatra que tenía un hijo hidranencefálico me contó un terrible dilema al que tuvo que enfrentarse siendo una joven madre. El neurocirujano que trataba a su bebé, antes de una operación para cerrarle la fontanela del cráneo, le sugirió que podía hacerse sin anestesia, puesto que el bebé carecía de corteza cerebral y, por tanto, no podía sentir dolor. No sé qué decisión tomaron al final ni tampoco me atreví a preguntárselo a mi colega. Sin embargo, casos así nos ayudan a ver hasta qué punto las consideraciones teóricas en apariencia abstractas pueden conducir en poco tiempo a espantosos errores médicos. De modo que en este punto seré tajante: si vamos a aceptar que alguien que parece ser consciente en realidad no lo es, deberíamos exigir un argumento sumamente convincente. No basta con plantear dudas filosóficas. Necesitamos un fundamento muy sólido para dar por hecho que en esas personas los dos tipos de conciencia se han separado, algo que en apariencia nunca ocurre en nosotros.

En estos momentos, ese fundamento lo proporciona la teoría cortical de la conciencia, según la cual la conciencia surge en exclusiva de la corteza. La aceptación de esta idea dentro de la neurociencia conductual es tan generalizada que a mí nunca se me ocurrió cuestionarla, hasta que, basándose precisamente en mi investigación sobre los sueños, Allen Braun me hizo prestar atención al misterioso papel que desempeña el tronco encefálico. Los médicos se basan en la teoría cortical para decidir si a determinados pacientes vivos hay que tratarlos o no como seres sintientes. Por ejemplo, hay cuidadores que crían a niños hidranencefálicos en condiciones de grave abandono

emocional, porque parten del supuesto de que son «vegetativos». Si la teoría es correcta, ningún problema; en el fondo no hay abandono porque esos niños son algo así como zombis filosóficos. Parecen tener sentimientos, pero en realidad no los tienen. El que parezca que los tienen es una ilusión producida por el problema de las otras mentes. Su comportamiento exterior nos engaña haciéndonos creer que tienen alguna naturaleza interior.[93]

Pero veamos qué es la teoría cortical de la conciencia y hasta qué punto es convincente.

Lo primero que hay que destacar es que empezó a desarrollarse muy pronto. La observación cotidiana de que nuestra conciencia se compone sobre todo de imágenes perceptivas de lo que ocurre a nuestro alrededor lleva a pensar que fluye a través de los sentidos. Sin duda, esta idea de sentido común nos ha acompañado desde que empezamos a plantearnos estas cuestiones, y en los siglos XVII y XVIII dio lugar a las filosofías «empiristas» de John Locke y David Hume, que en sus teorías sostenían que la mente —que al principio es una pizarra en blanco— va adquiriendo todas sus características específicas a partir de las impresiones que dejan las vibraciones sensoriales. Se suponía que esas impresiones acababan asociadas entre sí mediante conjunciones habituales diversas[94] para producir en nuestra memoria imágenes de los objetos, que a su vez servirían para construir sobre ellas ideas más abstractas. Las vibraciones sensoriales posteriores estimulan esas imágenes conjuntadas hacia el primer nivel de la conciencia, de modo que lo que experimentamos no son sensaciones en bruto, sino lo que hemos aprendido sobre el mundo.

A la forma en que las ideas se vuelven conscientes en respuesta a un estímulo externo se la llamó «apercepción» (que viene a significar percibir el presente a través de la lente de la experiencia pasada).[95] Procesos cognitivos como el uso del conjunto de imágenes mentales para pensar implicarían el mismo proceso a la inversa: activaciones generadas internamente de las imágenes guardadas en la memoria, convenientemente reordenadas (y más borrosas que las generadas externamente).

Pese a su carácter especulativo, los primeros neurólogos tomaron como punto de partida esta idea filosófica de la mente. Así, cuando en el siglo XIX los pioneros de la neuropsicología moderna intentaban determinar los marcadores neuronales de estos procesos, al observar que los órganos de los sentidos estaban conectados con la corteza

dedujeron que las «vibraciones» sensoriales tenían lugar en esos nervios conectores. No es que ignorasen el hecho de que los órganos de los sentidos están también conectados con los núcleos subcorticales, pero dieron por sentado que el gran almacén de imágenes de la memoria que constituye nuestro conocimiento del mundo tenía que hallarse en la corteza porque esta contiene muchísimas más neuronas. Así pues, la «apercepción» y las «ideas» asociadas que generan actividad mental propiamente dicha se consideraron fenómenos corticales. Theodor Meynert, la gran autoridad neuroanatómica de la época, lo expresó como sigue:

La función principal del órgano central es transmitir el hecho de la existencia a un yo que se va formando gradualmente en la corriente cerebral. […] Si consideramos la corteza como un órgano que funciona como un todo, lo único que se puede decir es que coadyuva los procesos de la mente. […] Pensar más sobre la corteza es imposible e innecesario.[96]

Por supuesto, la corteza cerebral humana tiene un tamaño impresionante.[97] El neuroanatomista y neurólogo Alfred Walter Campbell resumió así la opinión general al respecto durante la asamblea general anual de la Medico-Psychological Association celebrada en Londres en 1904:

Considerado colectivamente, el cerebro humano alberga dos clases de centros, que controlan lo que podemos llamar «funciones primarias» y «funciones evolutivas superiores», respectivamente; las primeras son las comunes a todos los animales y esenciales para la supervivencia, es decir, los centros para el movimiento y las sensaciones comunes y especiales; las segundas son aquellas funciones psíquicas complejas cuya posesión coloca al hombre por encima de todos los demás seres.[98]

Es importante reconocer que, desde este punto de vista, la mente está íntegramente constituida por imágenes de la memoria que reflejan experiencias pasadas del mundo exterior. Lo único que hacen las sensaciones entrantes es estimular esas imágenes y sus asociaciones hacia la conciencia. En consecuencia, las vibraciones que llegan

procedentes de los órganos de los sentidos son sucesos prementales —desencadenantes de la actividad mental—, pero no sucesos mentales entendidos como tales. Lo mismo ocurre con la función de los nervios que salen de la corteza y la conectan con el resto del cuerpo; esas vías no tienen nada de «mental», solo vierten los productos de la actividad mental. La actividad mental propiamente dicha solo puede tener lugar en la corteza, donde residen las imágenes de la memoria.

Meynert describió así la relación entre la corteza y el mundo exterior:[99]

Los efectos motores de nuestra conciencia que reaccionan ante el mundo exterior no son el resultado de fuerzas innatas del cerebro. El cerebro, como una estrella fija, no irradia su propio calor, y la energía subyacente a todos los fenómenos cerebrales la obtiene del mundo que hay fuera de él.

Creo que aquí convendría señalar que, como aún no había llegado Freud, se daba por hecho que toda actividad mental era consciente. Las palabras mental y consciente se utilizaban para designar lo mismo.

Todos los trabajos experimentales del siglo XIX que condujeron a nuestra concepción moderna de la conciencia se desarrollaron dentro de ese marco filosófico. A finales del siglo XIX, el fisiólogo Hermann Munk identificó la corteza occipital como el locus de la parte mental de la visión (véase fig. 6). Munk estudió el comportamiento de perros de laboratorio con lesiones infligidas en la corteza occipital. Estos desafortunados animales veían, pero al parecer no tenían un «entendimiento» normal de lo que veían: no podían reconocer visualmente a sus amos, por ejemplo, ni identificar sus propios comederos, aunque los miraban y daban vueltas a su alrededor y los reconocían mediante los demás sentidos.[100] Munk denominó a esta afección «ceguera mental», para distinguirla de la forma común de ceguera causada por lesiones de las vías sensoriales (subcorticales) que van de los ojos a la corteza.[101] Siguiendo la filosofía empirista, equiparó lo que llamó «visión mental» con la capacidad de activar imágenes de la memoria visual mediante asociaciones, en contraposición a la actividad mecánica de recibir sensaciones visuales en bruto y activar reflejos motores. Es lógico, pues, que el trastorno descrito por Munk se llame actualmente «agnosia» visual, esto es, falta de conocimiento visual.

Poco tiempo después se informó de fenómenos clínicos de la misma clase en seres humanos. En 1887, el oftalmólogo Hermann Wilbrand describió el caso de fräulein G., una mujer de sesenta y tres años que sufrió una apoplejía occipital bilateral:[102]

Todos los que la rodeaban la consideraban ciega, pero ella era bastante consciente de que no era completamente ciega, «porque cuando la gente se sentaba junto a mi cama y se compadecía de mi ceguera, yo pensaba: no debes de ser del todo ciega porque puedes ver el tapete que hay allí, de reborde azul, extendido sobre la mesa de la enfermería». […] Cuando esta señora —por lo demás muy inteligente — se levantó [tras una pérdida inicial de conocimiento debida a la apoplejía], se halló en un curioso estado de no ver y al mismo tiempo ver […]. Hoy todavía se emociona cuando recuerda su primera salida tras la apoplejía; lo distinta y extraña que le resultó la ciudad, y la angustia y agitación enormes que sintió cuando su acompañante la llevó por primera vez a pasear por el Jungfernstieg y el Neuer Wall hasta el Stadthaus y le iba presentando los edificios y las calles que tan familiares le resultaban antes. La paciente explica su reacción y sus respuestas: «Le dije a mi acompañante: "Si me dices que eso es el Jungfernstieg, eso el Neuer Wall y esto el ayuntamiento, pues supongo que tendrás razón, pero yo no los reconozco". […] Le dije a mi médico: "De mi estado se puede sacar la conclusión de que vemos más con el cerebro que con los ojos; el ojo no es más que el vehículo de la vista, porque lo veo todo con absoluta claridad y lucidez, pero no lo reconozco, y muchas veces ni sé qué puede ser lo que estoy viendo"».

Wilbrand llegó a la conclusión de que esta paciente no padecía una mera ceguera, sino una pérdida de memoria visual (un trastorno del conocimiento visual, el reconocimiento o el entendimiento), y señaló que, frente a los pacientes ciegos que pueden seguir generando imágenes visuales en sus sueños porque tienen intactas sus imágenes mentales, los pacientes con ceguera mental como fräulein G. no pueden hacerlo y pierden su capacidad de soñar. Lo que se preguntó Wilbrand fue cómo se podían generar alucinaciones visuales sin imágenes de la memoria visual.[103]

Más adelante, estas observaciones sobre la visión se generalizaron a otras modalidades de percepción. Volviendo a los perros, la ablación de la corteza auditiva les producía una «sordera mental», ahora llamada «agnosia auditiva», que les impedía responder a sonidos con

significado adquirido, aunque saltaba a la vista que no estaban sordos; respondían al ruido bruto, pero ya no reconocían sus nombres cuando los llamaban. En 1874, el neurólogo Carl Wernicke observó algo parecido en pacientes humanos y desarrolló el concepto de «afasia» para los trastornos del lenguaje adquirido.[104]

En la modalidad motora, Hugo Liepmann afirmó que las lesiones corticales también afectaban a la parte mental del movimiento y originaban la llamada «parálisis psíquica» (o «apraxia»). Una lesión de las vías motoras de salida provocaba parálisis física, pero una lesión en el centro cortical de las imágenes de la memoria motora provocaba el olvido de habilidades de movimiento adquiridas («apraxia motora»), en tanto que una lesión de las vías transcorticales de las asociaciones motoras provocaba una desconexión entre el movimiento hábil y las ideas abstractas, como el significado simbólico de los gestos de la mano («apraxia ideomotora»).[105]

A partir de ahí, la ceguera, la sordera y la parálisis subcorticales se consideraron trastornos físicos, y los efectos visuales, auditivos y motores de las lesiones corticales, trastornos mentales, es decir, las formas aperceptivas y asociativas de la agnosia, la afasia y la apraxia. A estos trastornos se los llama ahora trastornos «neurocognitivos». Obsérvese que la distinción entre los dos tipos de trastornos (subcorticales frente a corticales) coincidía así con la separación disciplinaria entre la neurología y la neuropsicología, como sigue ocurriendo.

Cuando los contemporáneos de Munk les extirparon a los perros toda la corteza —en lugar de solo las partes especializadas responsables de las distintas modalidades sensoriomotoras—, los animales no entraron en coma ni en estado vegetativo, sino que se comportaron como si estuvieran «sin mente» (según entendían la expresión los empiristas). Se volvieron amnésicos, esto es, perdieron todas sus imágenes de la memoria y, por tanto, su «entendimiento», lo que llevó a Friedrich Goltz a calificarlos de «idiotas». A aquellos primeros investigadores no les sorprendió que los animales no entraran en coma porque, partiendo de los supuestos teóricos de la época, lo único que esperaban era que los perros decorticados perdieran todos los conocimientos adquiridos. La pérdida no tenía por qué afectar a las sensaciones corporales ni a los reflejos, que se suponían subcorticales. Por aquel entonces nadie discutía la idea de que la vida mental solo consistía en imágenes de la memoria. Las controversias de la época giraban en torno a otras cosas, como la precisión con la que podía ubicarse cada función mental en regiones concretas de la corteza, pero nadie dudaba de que todas aquellas funciones fueran corticales.

Tanto es así que, cuando resumió todo el panorama emergente en su famoso libro de 1884 Psychiatrie: Klinik der Erkrankungen des Vorderhirns,[106] Meynert identificó la mente con la totalidad de las imágenes de la memoria de los objetos producidas por la proyección de la periferia sensoriomotora sobre la corteza, más las asociaciones transcorticales entre ellas y las imágenes de la memoria que constituían ideas abstractas. Naturalmente, situó todas esas imágenes, asociaciones e ideas en la corteza, como confirma la cita que hemos visto anteriormente. La llamaba la parte «voluntaria» del encéfalo y afirmaba que tenía conexiones directas con la periferia corporal a través de los nervios sensoriales y motores. En otras palabras, afirmaba que la corteza estaba conectada con el cuerpo independientemente de la materia gris subcortical. Reconoció que la parte subcortical del encéfalo también estaba conectada con la corteza y la periferia corporal a través de sus propias vías independientes, pero a estas las calificó de «reflejas». Estamos ante los «dos cerebros» de los que hablaba antes para explicar el comportamiento de los niños hidranencefálicos.

De esta forma, la corteza pasó a ser el órgano de la mente —entendida como conciencia de las imágenes de la memoria—, y el cerebro subcortical se quedó sin mente. Todo se redujo a la idea de que nuestra «mente» está constituida en su totalidad por experiencias pasadas, que dejan huellas que se asocian entre sí para formar imágenes concretas e ideas abstractas. El resto de nosotros, las partes sensoriomotoras periféricas y las partes innatas y subcorticales, incluidas las que transmiten impresiones del interior de nuestro cuerpo, se consideraban puro reflejo. Y así, por extraño que parezca, la distinción filosófica entre nuestra mente y nuestro cuerpo llegó a coincidir con la distinción anatómica entre la corteza y la subcorteza.

En la década de 1960, Norman Geschwind, el gran pionero de la neurología conductista en Estados Unidos, rescató con entusiasmo estas ideas clásicas,[107] y ese renacimiento del asociacionismo empirista en neurología coincidió con la «revolución cognitiva» en psicología. Los diagramas de flujo de información de los científicos cognitivos se parecían mucho a los diagramas que los neurólogos alemanes clásicos habían utilizado para ilustrar las relaciones funcionales entre los centros corticales que contenían imágenes de la memoria de diversos tipos. Sin embargo, como veremos más adelante, el cognitivismo moderno también llevó a la constatación —que acabó de consolidarse en la década de 1990— de que, después de todo, muchas funciones mentales (incluidas la percepción y la memoria) no eran conscientes.

Casi al mismo tiempo que Geschwind resucitaba ese vínculo entre la neuropsicología y la teoría cortical clásica, los pioneros de lo que más tarde se convertiría en la neurociencia afectiva (Paul Maclean y James Olds, entre otros) acumulaban observaciones que revelaban que muchos núcleos subcorticales que conectan el cerebro con el interior del cuerpo también realizaban funciones mentales, como la pulsión motivacional y el sentimiento emocional.

Pese a estos desarrollos que enseguida comentaré, paralelos a la neurociencia cognitiva y a la afectiva, obsérvese que esta herencia intelectual —el intento de los neurólogos del siglo XIX de confirmar las teorías de los filósofos del siglo XVIII— es la que en última instancia fundamenta el supuesto de los neurólogos actuales de que los niños hidranencefálicos no tienen mente, algo que dan por sentado incluso cuando reconocen que las antiguas ecuaciones entre «mente», conciencia y corteza perdieron su validez: que no toda la actividad mental es consciente y no toda la conciencia es cortical. En otras palabras, la insistencia en que la ecuación original sigue en pie, que la «conciencia como experiencia» es necesariamente cortical, se basa en la inercia teórica y no en la evidencia científica.

Llegados a este punto, ya habrán deducido que a mí la teoría cortical de la conciencia no me parece válida. De hecho, incluso me atrevería a afirmar que los animales y los seres humanos pueden estar en estado consciente incluso cuando carecen por completo de corteza. Creo que sí hay «algo que es como» ser una niña hidranencefálica.

Ciertamente, una de las razones por las que cuesta tanto saber qué experimentan los pacientes hidranencefálicos —si es que llegan a tener experiencia interior— es que no pueden hablar. Dado que el lenguaje es, sin duda, una función cortical, no podemos esperar que las personas que carecen de corteza nos proporcionen informes verbales introspectivos. Esas personas no nos pueden dar la misma evidencia subjetiva con la que nos convencemos en otros casos del estado consciente de alguien, con independencia del problema de las otras mentes.[108] Y con los animales ocurre igual. Sin embargo, hay seres humanos que, habiendo perdido una parte considerable de la corteza cerebral, conservan la capacidad de hablar, y en estos casos basta con preguntarles qué experimentan.

Tras extirpar, bajo anestesia local, grandes secciones de corteza (incluso hemisferios enteros) en setecientos cincuenta pacientes humanos, sobre todo para el tratamiento de la epilepsia, los

neurocirujanos Wilder Penfield y Herbert Jasper observaron que este tipo de intervención tiene efectos limitados en la conciencia autoinformada, incluso en el momento mismo en que se extirpa la corteza. (La cirugía cerebral suele realizarse bajo anestesia local para que el paciente pueda informar sobre los efectos de cada acción del cirujano). Penfield y Jasper llegaron a la conclusión de que las resecciones corticales no interrumpen el ser sintiente; simplemente privan a los pacientes de «ciertas formas de información».[109] Yo mismo he asistido a numerosas operaciones de este tipo, en las que mi función es evaluar los efectos de la estimulación eléctrica de las partes de la corteza de memoria y lenguaje antes de que el cirujano empiece a cortar. Y he sido testigo de lo que Penfield y Jasper informan.

Sin embargo, en la corteza no todo es igual. Las regiones que reciben los aportes de cada uno de nuestros sentidos especiales son las que generan los qualia fenoménicos asociados a dichos sentidos (color para la visión, tono para la audición, entre otros). En cambio, según la opinión establecida, para poder informar de las sensaciones en bruto es necesario que acceda a ellas un tipo de conciencia dominante que constituye el «yo» sintiente.[110]

Para describir la relación entre la conciencia «fenoménica» (por ejemplo, ver y oír sin más) y la conciencia «reflexiva» (saber que estás viendo y oyendo) se emplean distintas terminologías, pero todas transmiten la misma idea básica: somos algo más que las diversas formas de información sensorial que procesamos. Por eso hay pacientes que, teniendo lesiones en la corteza visual primaria o en la corteza auditiva, pueden ser ciegos o sordos, respectivamente, pero mantienen su sentido del yo. Este yo sintiente es el que adivina lo que uno está viendo cuando está desprovisto de qualia visuales por tener ceguera, por ejemplo. Esos pacientes solo pierden «ciertas formas de información». Surge entonces la gran pregunta: ¿qué podemos aprender de los pacientes que han perdido esas partes de la corteza que se supone que son responsables de la mismidad?

Hay tres ubicaciones candidatas para esta función. La primera es la corteza insular, especializada en la conciencia interoceptiva y que, según el consenso general, genera los sentimientos que constituyen un «yo» sintiente.[111] La segunda es la corteza prefrontal dorsolateral, que forma una superestructura sobre todas las demás partes del encéfalo y a la que la mayoría atribuye los «pensamientos de orden superior», incluida la conciencia de las sensaciones y los sentimientos.[112] La tercera es la corteza anterior del cíngulo, que se ilumina en casi todos los experimentos de técnicas de imagen sobre el cerebro cognitivo y que se supone que media en aspectos como el

«procesamiento relacionado con el yo» y la «voluntad».[113]

Veamos estas tres regiones una por una.

Lo que sigue es un fragmento de una entrevista realizada por Antonio Damasio a un paciente cuya corteza insular quedó totalmente destruida a causa de una enfermedad vírica llamada encefalitis por herpes simple:[114]

P: ¿Tiene usted sentido del yo?

R: Sí.

P: ¿Y si le dijera que ahora mismo no está aquí?

R: Le diría que está usted ciego y sordo.

P: ¿Cree que otras personas pueden controlar sus pensamientos?

R: No.

P: ¿Y por qué cree que eso no es posible?

R: Quiero creer que cada cual controla su mente.

P: ¿Y si le dijera que su mente era la mente de otra persona?

R: ¿Y cuándo se hizo el trasplante? Me refiero al trasplante cerebral.

P: ¿Y si le dijera que lo conozco mejor de lo que usted se conoce a sí mismo?

R: Pensaría que se equivoca.

P: ¿Y si le dijera que usted es consciente de que soy consciente?

R: Le daría la razón.

P: ¿Es usted consciente de que yo soy consciente?

R: Soy consciente de que usted es consciente de que yo soy consciente.

El equipo de Damasio estudió a este paciente, llamado «B.», durante veintisiete años seguidos (adquirió la enfermedad con cuarenta y ocho

años y murió a los setenta y tres). En el estudio de caso publicado expusieron numerosos datos neuropsicológicos relativos a su capacidad de sentir y responder emocionalmente, y llegaron a las siguientes conclusiones:

El paciente B., que tenía las cortezas insulares totalmente destruidas, experimentaba sensaciones corporales y sentimientos. Informaba de que sentía dolor, placer, picor, cosquillas, felicidad, tristeza, aprensión, irritación, cariño y compasión, y cuando experimentaba alguna de esas sensaciones o sentimientos se comportaba en consecuencia. También refería sensaciones de hambre, sed y ganas de evacuar, y también en esos casos se comportaba en consecuencia. Deseaba tener oportunidades de jugar (por ejemplo, a las damas), de charlar con alguien o de salir a pasear, y cuando participaba en alguna de esas actividades daba muestras claras de placer, como las daba también de decepción o incluso irritación cuando se le negaban esas oportunidades. [...] Dado el empobrecimiento de su imaginación, la existencia del paciente B. era una reacción «afectiva» prácticamente constante a sus propios estados corporales y a las modestas exigencias planteadas por el mundo que lo rodeaba, sin el filtro de controles cognitivos de orden superior. [...] Según Craig (2011), el signo revelador de la autoconciencia es la capacidad de reconocerse en un espejo, una capacidad que, en sus palabras, «solo puede proporcionar una representación neuronal funcional y emocionalmente válida de uno mismo». El paciente B. superó esta prueba una y otra vez. En resumen, estos hallazgos desmentirían la idea de que la autoconciencia humana, junto con la capacidad de sentir, dependen por completo de las cortezas insulares y, en concreto, de su tercio anterior (Craig, 2009, 2011). En ausencia de cortezas insulares, debemos considerar alternativas neuroanatómicas para explicar de dónde surgen las capacidades de sentir y la sintiencia del paciente B. [115]

Bud Craig, entre otros, dice que las sensaciones afectivas que constituyen el yo se generan en la corteza insular, pero Damasio descubrió que el paciente B. era, de hecho —al igual que las ratas decorticadas—, más emocional después de sufrir la lesión cortical que antes. La predicción de que tales pacientes perderían su «presencia» sintiente queda claramente desmentida.

Lo mismo ocurre con los pacientes humanos que han sufrido lesiones

en la segunda región que nos incumbe, la corteza prefrontal, que forma una superestructura sobre todas las demás partes del encéfalo. La emocionalidad de los pacientes prefrontales es bien conocida porque se trata de una de las principales características del «síndrome del lóbulo frontal», algo que ya se observó en 1868 en el célebre caso de Phineas Gage, a quien una barra de hierro le atravesó el cráneo en un accidente laboral.

El equilibrio entre sus facultades intelectuales y sus instintos animales parece haber sido destruido. Es inestable, irreverente, entregándose en ocasiones a la blasfemia más grosera (cosa que no hacía antes), manifestando muy poco respeto por sus compañeros, incapaz de contenerse cuando entra en conflicto con sus deseos.[116]

Algunos de los neurocientíficos de la emoción más importantes, como Joseph LeDoux, creen que las sensaciones y los sentimientos nacen realmente en la corteza prefrontal dorsolateral.[117] Según esta teoría, los precursores subcorticales de las sensaciones y los sentimientos son totalmente inconscientes hasta que son «etiquetados» en la memoria de trabajo consciente.[118] Para estos teóricos, la emoción solo es otra forma de cognición; de hecho, una forma de cognición bastante abstracta y reflexiva. Pero, si tienen razón, ¿por qué los pacientes con la mayor parte de su memoria de trabajo destruida manifiestan tantas emociones? Phineas Gage solo es el caso más famoso de muchos otros recogidos en la bibliografía. ¿Dónde queda aquí el papel de la conciencia reflexiva cuando la forma prerreflexiva del sentimiento se expresa de una forma tan vívida en el comportamiento de estos pacientes?

Para ser sinceros, a la mayoría de los teóricos corticales no les interesan mucho las sensaciones y los sentimientos. Se centran más bien en el modo en que contribuyen los lóbulos prefrontales al pensamiento de orden superior. Por otra parte, las hipótesis que se derivan de estas teorías más amplias se pueden probar fácilmente en casos humanos con lesiones en los lóbulos prefrontales. Las lesiones prefrontales completas son muy raras, pero se conoce su existencia.

Mi paciente W. tenía cuarenta y ocho años cuando lo examiné por primera vez. Obtuvo entonces la máxima puntuación en la escala de coma de Glasgow, lo que significa que estaba totalmente despierto y reactivo. A los trece años sufrió la rotura de un aneurisma cerebral y

tuvieron que operarlo. En la intervención se retrajeron los lóbulos frontales para envolver el vaso sanguíneo vulnerable y evitar así nuevas hemorragias. Lamentablemente, la operación causó una infección cerebral crónica que requirió muchas otras intervenciones, hasta llegar a la destrucción total de los lóbulos prefrontales de ambos lados (véase fig. 7, p. 92).

Por fortuna, su corteza de lenguaje se mantuvo intacta. Reproduzco a continuación parte de nuestra conversación:

P: ¿Es usted consciente de sus pensamientos?

R: Sí, claro.

P: Para confirmarlo, voy a pedirle que resuelva un problema para el que es necesario que se imagine una situación de forma consciente.

R: Vale.

P: Imagine que tiene dos perros y una gallina.

R: Vale.

P: ¿Los está viendo mentalmente?

R: Sí.

P: Pues dígame por favor cuántas patas ve en total.

R: Ocho.

P: ¿Ocho?

R: Sí; los perros se comieron al pollo.

El paciente W. dijo esta última frase con una sonrisa traviesa. Puede que no sea el mejor chiste del mundo, pero constituye una prueba convincente de que, por decirlo coloquialmente, «había alguien en casa». El paciente W. afirmó que estaba consciente, como había hecho el paciente B. de Damasio, y, a falta de una razón más convincente que el escepticismo filosófico radical, yo me inclino a creerlo.

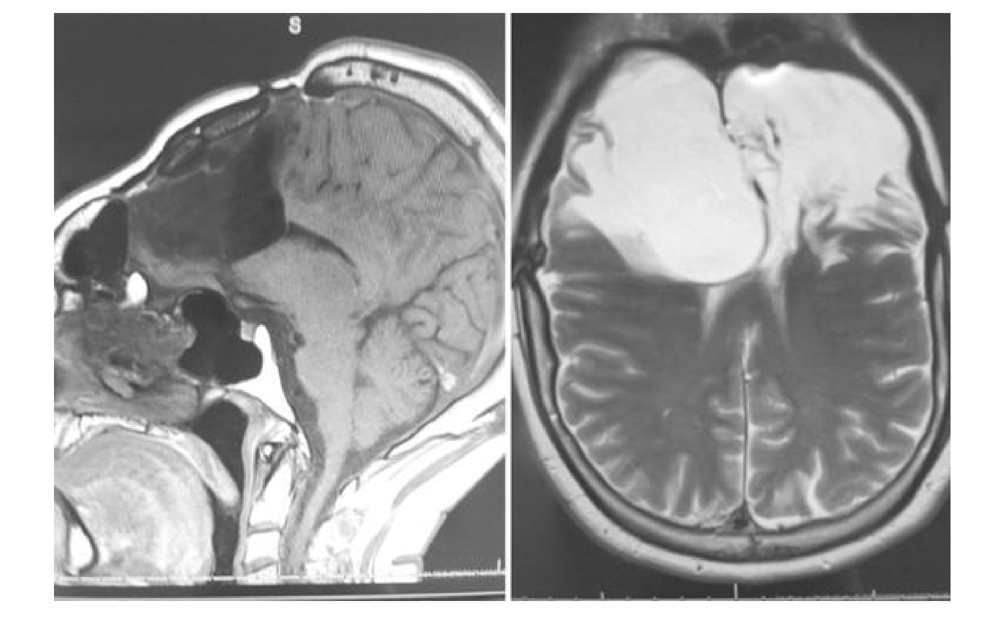


Figura 7

. Resonancia magnética cerebral del paciente W., que muestra la destrucción completa bilateral de los lóbulos prefrontales.

La tercera y última región cortical que se supone que guarda especial relación con el yo sintiente es la circunvolución del cíngulo anterior. Es relativamente fácil encontrar pacientes con lesiones bilaterales completas en esa región, entre otras razones porque esa parte de la corteza se operaba a menudo para tratar enfermedades psiquiátricas, como el trastorno obsesivo-compulsivo. (Precisamente la asociación de la circunvolución del cíngulo anterior con el «procesamiento relacionado con el yo» es lo que ha hecho que se fijen tanto en ella).

En la fase posoperatoria aguda, algunos de estos pacientes sufren una crisis en la distinción entre fantasía y realidad. Charles Whitty y Walpole Lewin describieron un ejemplo sorprendente (su caso 1).

Cuando le preguntaron qué había hecho durante el día, el paciente respondió: «He tomado el té con mi mujer». Antes de que le dijeran nada más, añadió: «Bueno, en realidad, no. Hoy no ha venido. Pero, en cuanto cierro los ojos, se produce esa escena con claridad. Yo veo las tazas con sus platos y la oigo servir el té. Entonces, en el momento

en el que levanto la taza para beber, yo me despierto y no hay nadie».

P: Así pues, ¿duerme usted mucho?

R: No, yo no estoy dormido, es como un sueño despierto […] a veces incluso con los ojos abiertos […]. Es como si mis pensamientos estuvieran descontrolados y tuvieran vida propia, ¡todo parece tan real! La mitad de las veces yo no estoy seguro de si solo lo he pensado o realmente ha pasado.[119]

Obsérvese el frecuente uso del «yo». Whitty y Lewin sospecharon que su paciente podría haber sufrido crisis complejas, pero se han descrito experiencias similares en muchos otros casos con lesiones bilaterales de la circunvolución del cíngulo anterior.

El ejemplo siguiente es un caso mío.[120] Se trata de una mujer de cuarenta y cuatro años que había sufrido una hemorragia cerebral subaracnoidea. Este es su relato:

Es como si mi pensamiento se hiciera realidad; como si, solo por pensar en algo, viera cómo ocurre ante mis ojos, y entonces me siento muy confundida y no sé qué es lo que ha pasado de verdad y qué es lo que estoy pensando.

Mi paciente puso un ejemplo:

Paciente: Estaba tumbada en la cama, pensando, y de pronto, surgido de la nada, tenía delante a mi marido [fallecido], hablándome. Y luego fui a bañar a los niños y, de repente, abrí los ojos y «¿Dónde estoy?». ¡Y estaba sola!

Yo: ¿Se había quedado dormida?

Paciente: No creo; es como si mis pensamientos se hubieran hecho realidad.

Es evidente que en estos casos la conciencia está alterada, pero esa no

es la cuestión. Nadie discute la implicación de la corteza en el procesamiento consciente. Lo que es insostenible es la idea de que el yo consciente se genera ahí.

Hemos visto comportamientos animados, deliberados, reactivos y emocionales en humanos y animales que carecen por completo de corteza cerebral. La introspección sugiere que la conciencia contribuye en gran medida a dichos comportamientos en nuestro propio caso (presuntamente normal). Por lo tanto, con independencia del problema de las otras mentes, haría falta una razón de peso para pensar que no ocurre lo mismo en esos casos. Sin embargo, cuando buscamos esas razones de peso, lo único que encontramos es la carga de la historia académica, un modelo preneurológico de la mente que utilizaban como plantilla los primeros exploradores del cerebro. En la actualidad, muy pocos científicos respaldarían la versión de Hume del empirismo o la Psiquiatría de Meynert. De hecho, a las teorías de Meynert se las ha llegado a llamar «mitología cerebral».[121] Y, con todo, el dogma de que la corteza es el órgano de la mente se ha convertido en el supuesto fundamental de todo un campo de la medicina.

Ciertamente, la corteza está implicada en muchas funciones cognitivas, como el lenguaje, lo cual significa que no podemos esperar informes verbales introspectivos de las personas que carecen por completo de ella. Así y todo, tenemos testimonios en primera persona de pacientes cuyos trastornos les permiten declaraciones verbales autorreflexivas, incluso careciendo de las partes concretas de la corteza que en teoría dan lugar a una conciencia global. Esos pacientes afirman, una y otra vez, estar conscientes y exponen su «ser» introspectivo. ¿Acaso mienten? Si así fuera, ¿por qué? ¿Y cómo se entendería que una persona que no tiene yo mienta sobre su mismidad? Aquí la neurociencia cognitiva bordea la incoherencia, una buena señal de que ha tomado el camino equivocado.

Ateniéndome a las pruebas, mi opinión es que la teoría cortical no se sostiene. No hay ninguna buena razón para creer que la corteza da lugar a una existencia sintiente tal y como ustedes y yo la experimentamos de ordinario, mientras que sí hay muchas buenas razones para concluir lo contrario. Tendremos que buscar el manantial de nuestro ser en otra parte.

[82] Merker, 2007, p. 79. Estas observaciones confirmaron un informe anterior de Shewmon, Holmes y Byrne, 1999.

[83] La siguiente descripción de los hallazgos de Merker parafrasea su informe publicado en 2007.

[84] Damasio y Carvalho, 2013, p. 147.

[85] Estas observaciones parafrasean el resumen que hace Merker de lo publicado al respecto, 2007, p. 74.

[86] Panksepp, 1998.

[87] Véase Weiskrantz, 2009.

[88] Véase el vídeo: https://blogs.scientificamerican.com/observations/blindsight-seeing-without-knowing-it/.

[89] Lo mismo ocurre con la vía que va del ojo al cuerpo geniculado lateral. Véase fig. 6.

[90] Zeman, 2001.

[91] Coenen, 2007, p. 88; la cursiva es mía.

[92] Ausencia de apertura ocular = 1 punto, apertura en respuesta al dolor = 2 puntos, apertura en respuesta al habla = 3, apertura espontánea = 4; ausencia de respuesta verbal = 1, respuesta con sonidos incomprensibles = 2, respuesta con palabras inadecuadas = 3, respuesta coherente pero inadecuada = 4, respuesta adecuada = 5; ausencia de respuesta motora = 1, postura de descerebración = 2, postura de decorticación = 3, alejamiento del dolor = 4, localización del dolor = 5, obediencia a órdenes = 6.

[93] También se ha afirmado que esa ilusión se debe a la «falacia moralista». La neurocientífica cognitiva Heather Berlin dijo que era «arbitrario» suponer que la capacidad de respuesta externa implica conciencia interna, y lo elaboró como sigue (Berlin, 2013, pp. 25-26): «La hipótesis principal de Solms de que los niños hidranencefálicos son conscientes es injustificada. No podemos dar por hecho que para tener un ciclo de sueño-vigilia y expresiones de emoción (risa, rabia, etc.) haga falta conciencia. […] Si bien es cierto que pueden llegar a ser conscientes, no por eso podemos dar por hecho que lo sean. Los procesos inconscientes pueden ser bastante sofisticados y complejos

(Berlin, 2011). El quid de la teoría de Solms se basa en una proyección de la existencia de conciencia a partir de lo que parecen comportamientos emocionales significativos, un ejemplo de la "falacia moralista" (argumentar que algo tiene que ser cierto porque creerlo así nos haría sentir bien). Los seres humanos tienen un deseo natural de dar por hecho que la conciencia existe».

Esta fue mi respuesta (Solms, 2013, pp. 80-81): «¿Por qué deberíamos dar por hecho que unas manifestaciones emocionales contextualmente apropiadas —provocadas por la estimulación de una región cerebral concreta y suprimidas por lesiones en esa misma región cerebral— que en nosotros corresponden a sentimientos afectivos no corresponden a sentimientos afectivos en estos niños y animales? Sin duda esa suposición sería más «arbitraria» que la mía. Porque la única prueba es que esos niños y esos animales no pueden «declarar» sus sentimientos con palabras».

[94] Conocidas como las «leyes de asociación»: por contigüidad, repetición, atención, semejanza, *etc.*

[95] La apercepción es «el proceso por el cual la nueva experiencia es asimilada y transformada por el residuo de la experiencia pasada de un individuo para formar un nuevo todo» (Runes, 1972).

[96] Meynert, 1867.

[97] Sin tener en cuenta, claro está, que es más grande en otros mamíferos (como los elefantes). De hecho, la corteza humana ni siquiera es mayor que la de otros mamíferos si consideramos la relación entre ella y el tamaño corporal, o entre la corteza y la subcorteza.

[98] Campbell, 1904, pp. 651-652.

[99] Meynert, 1884 (trad. inglesa, 1885), p. 160.

[100] Munk, 1878, 1881.

[101] La base anatómica de la distinción entre ceguera y ceguera mental se atribuía al hallazgo de Paul Flechsig (1901, 1905), a saber, que la corteza estriada «de proyección» contiene células primordiales que están conectadas directamente con la periferia retiniana. Estas células primordiales están mielinizadas al nacer y, por tanto, no contienen imágenes de la memoria. La corteza circundante «de asociación» —el vehículo de todas las funciones mentales— se mieliniza mucho más tarde. Lo mismo ocurre con las otras cortezas

específicas de otras modalidades.

[102] Wilbrand, 1887, 1892. Véase mi traducción al inglés del caso clínico original de Wilbrand, donde se comentan en detalle muchos de estos aspectos teóricos (Solms, Kaplan-Solms y Brown, 1996).

[103] Esta observación de Wilbrand se basaba en una observación similar anterior de Charcot (1883) sobre un paciente que experimentaba sueños no visuales. De ahí derivó el concepto de «síndrome de Charcot-Wilbrand»: incapacidad para volver a visualizar y reconocer objetos visuales durante el día y pérdida de sueños durante la noche. Para un análisis crítico del concepto Charcot-Wilbrand a la luz de mis hallazgos posteriores, véanse Solms, Kaplan-Solms y Brown, 1996; y Solms, 1997a.

[104] Wernicke marcó una distinción entre la sordera común y la «sordera de palabras» (afasia). La afasia de Wernicke se generaba por lesiones en el área cortical que contiene las imágenes de la memoria auditiva de las palabras —recuerdos de los sonidos del habla—, en tanto que la sordera común se debía a lesiones en las vías subcorticales que conectan esta área con las sensaciones auditivas entrantes. Años antes, Paul Pierre Broca (1861, 1865) había descrito una forma paralela de afasia, causada por lesiones en las imágenes motoras de las palabras: lesiones en los programas aprendidos sobre cómo producir los sonidos del habla. Ludwig Lichtheim (1885), alumno de Wernicke, aún añadió otras formas de afasia debidas a lesiones en las vías transcorticales «de asociación» que llevaban de las imágenes de las memorias auditivas y motoras para las palabras a las ideas abstractas, que daban significado a las imágenes concretas.

Heinrich Lissauer (1890), por su parte, subdividió la «ceguera mental» en dos tipos: uno «aperceptivo», causado por lesiones en las propias imágenes de la memoria visual, y otro «asociativo», causado por lesiones en las vías transcorticales que van desde las imágenes de la memoria de los objetos visuales hasta las imágenes de las ideas abstractas. Fue Freud (1891) quien más tarde rebautizó la ceguera mental y la llamó «agnosia» visual.

Posteriormente, Hugo Liepmann (1900) también dividió la «parálisis psíquica» (apraxia) en tipos aperceptivos y asociativos: «motora» e «ideomotora», respectivamente, como se recoge en el texto.

[105] Liepmann, 1900.

[106] Meynert escribió en el prefacio: «El lector no encontrará en este

libro otra definición de “psiquiatría” que la que figura en la portada: Tratado clínico de las enfermedades del prosencéfalo. El término histórico de la psiquiatría, esto es, “tratamiento del alma”, implica más de lo que podemos realizar y trasciende los límites de la investigación científica exacta».

[107] Véanse Absher y Benson, 1993; y Goodglass, 1986. Curiosamente, Antonio Damasio fue alumno de Geschwind.

[108] Por otro lado, sabemos que los pacientes corticales sin lenguaje conservan plena conciencia, porque pueden comunicar sus sentimientos de otras formas, y de hecho lo hacen. Véase Kaplan-Solms y Solms, 2000, para descripciones detalladas de informes introspectivos no verbales de pacientes afásicos de diversos tipos. Superadas ciertas dudas iniciales en el siglo XIX, siempre ha habido consenso general en lo relativo a que la pérdida del lenguaje no afecta significativamente a la «inteligencia».

[109] Merker, 2007, p. 65.

[110] Véase la distinción de Ned Block (1995) entre «conciencia fenoménica» y «conciencia de acceso».

[111] Véanse Craig, 2009, 2011.

[112] Véanse Dehaene y Changeux, 2005; Baars, 1989, 1997. La expresión «pensamiento de orden superior» procede de Rosenthal, 2005.

[113] Qin et al., 2010; Mulert et al., 2005.

[114] El siguiente diálogo está extraído de un informe oral del caso presentado durante el Congreso de Neuropsicoanálisis de Berlín, 2011. El estudio de caso se publicó después en Damasio, Damasio y Tranel, 2013.

[115] Damasio, Damasio y Tranel, 2013.

[116] Harlow, 1868.

[117] LeDoux y Brown, 2017.

[118] Véase LeDoux, 1999, p. 46 (la cursiva es mía): «Cuando los estímulos eléctricos aplicados a la amígdala de los seres humanos provocan sentimientos de miedo (véase Gloor, 1992), no es porque la amígdala “sienta” miedo, sino porque en el fondo las distintas redes

que la amígdala activa envían a la memoria de trabajo estímulos etiquetados como miedo».

[119] Whitty y Lewin, 1957, p. 73.

[120] Solms, 1997a, p. 186, caso 22.

[121] El epíteto lo introdujo Karl Jaspers (1963).

04

¿Qué experimentamos?

¿Hay que ser consciente de lo que se percibe y se aprende para percibirlo y aprenderlo? El sentido común quizá dirá que sí. Los filósofos empiristas decían que sí. Pero la respuesta es que no.

Las ideas de los empiristas dieron lugar a la opinión clásica neuroanatómica según la cual la percepción (y las huellas mnémicas que de ella se derivan) es el ingrediente básico de la conciencia. Sin embargo, la evidencia científica actual sugiere de forma incontestable que somos inconscientes de la mayor parte de lo que percibimos y aprendemos. La percepción y la memoria no son funciones cerebrales inherentemente conscientes. Aquí el sentido común se equivocaría, porque resulta que todo lo que hace nuestra mente —con una sola excepción que veremos más adelante— se puede hacer bastante bien de manera inconsciente.

Lo más sorprendente es que esta reflexión es atribuible a Sigmund Freud.

Aunque durante toda su vida apoyó la teoría cortical de la conciencia, Freud fue uno de los primeros neurocientíficos que cuestionaron la distinción que hacía su maestro Meynert entre las partes «mentales» y «no mentales» del cerebro.[122] En 1891, tras examinar los datos entonces disponibles, Freud llegó a la conclusión de que las ideas de «Munk y otros investigadores que se basan en Meynert [...] ya no se sostienen».[123] En concreto, no encontró ninguna distinción anatómica clara entre las vías «voluntarias» y las vías «reflejas» de Meynert. Asimismo, demostró (contando el número de fibras nerviosas implicadas) que las imágenes de la memoria cortical solo se generaban después de una serie de conexiones intermedias entre ellas y la periferia sensorial. Dichas conexiones se hallaban en las partes supuestamente no mentales del encéfalo y reducían la cantidad de información que se transmitía en cada etapa. Freud infirió entonces que esas conexiones subcorticales debían de hacerle algo a la información sensorial que procesan. Observen su enigmática metáfora, concebida en una época predigital:

Solo sabemos que las fibras que llegan a la corteza cerebral tras su

paso por los tejidos grises [subcorticales] siguen manteniendo alguna relación con la periferia del cuerpo, pero ya no pueden ofrecer una imagen que se le parezca topológicamente. Contienen la periferia corporal de la misma forma en la que […] un poema contiene el alfabeto, totalmente reorganizado, para servir a otros fines, con conexiones múltiples entre los distintos elementos topológicos, por lo que algunos pueden ser representados varias veces y otros, ninguna. [124]

Si la corteza no está conectada con la periferia corporal de forma directa sino a través de conexiones subcorticales intermedias, entonces las imágenes de la memoria depositadas en la corteza no pueden ser proyecciones literales del mundo exterior. Tienen que ser el producto final de un procesamiento de información de varias etapas. Y, dado que dicho procesamiento culmina en imágenes de la memoria cortical, la parte subcortical del procesamiento tiene que generar de alguna forma versiones preliminares de las imágenes de la memoria. En consecuencia, las conexiones subcorticales deben aportar parte del procesamiento mental que llamamos «apercepción». No tiene sentido, decía Freud, trazar una línea artificial entre las partes subcorticales y las partes corticales del procesamiento y afirmar que solo el producto final es «mental»:

¿Acaso tiene sentido coger una fibra nerviosa —que en todo su recorrido se ha limitado a ser una estructura fisiológica sujeta además a modificaciones fisiológicas— y sumergir su extremo en la mente y revestirlo con una idea o con una imagen de la memoria?[125]

Al suponer —como lo hicieron todos los demás en aquella época— que solo son conscientes los procesos corticales, Freud albergó la idea de que la percepción y el aprendizaje tienen que incluir etapas preliminares inconscientes que son igual de «mentales» que las conscientes. Es decir, las huellas mnémicas inconscientes (subcorticales) tienen que ser tan mentales como las corticales, porque también forman parte de la función que llamamos «memoria», aunque carezcan de conciencia. En una carta a Wilhelm Fliess, Freud expuso su conclusión: «La novedad esencial de mi teoría radica en la tesis de que la memoria no está presente una, sino múltiples veces, que está registrada en varios tipos de signos».[126]

Freud identificó cinco etapas sucesivas en el procesamiento de la información: percepción, huella perceptiva, inconsciente, preconsciente y consciente (véase fig. 8, p. 98). La diferencia fundamental entre estas etapas no radicaba en que el inconsciente era corporal y el preconsciente mental, sino en que el tipo preconsciente de procesamiento de memoria se podía reproducir en la conciencia, en tanto que el tipo inconsciente no.[127] Es decir, solo una parte de la mente es consciente.

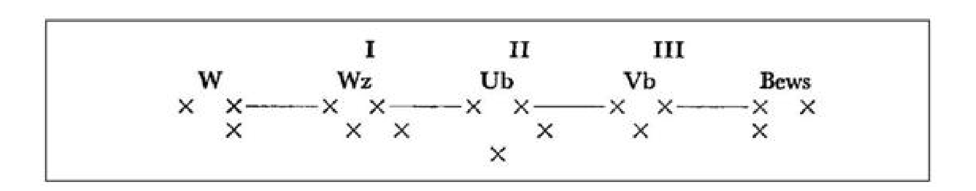


Figura 8

. Primer diagrama de Freud de los sistemas de memoria, que acompañaba a su carta a Wilhelm Fliess. (W = percepción, Wz = huella perceptiva, Ub = inconsciente, Vb = preconsciente, Bews = consciente).

De forma similar, en 1895 Freud habló sobre las vías nerviosas que proceden del interior del cuerpo. Tampoco en este caso, decía, tiene sentido afirmar que la información sobre los estados corporales solo se vuelve «mental» cuando llega a la corteza. Puesto que esta información —que transmite sensaciones como el hambre y la sed— impone demandas de trabajo a la mente a consecuencia de su conexión con el cuerpo, también tendremos que dar cabida a esas demandas en nuestra imagen de la mente. Aunque no seamos conscientes de nuestras «pulsiones» biológicas —como las llamaba Freud—, es indudable que forman parte de nuestra mente. Recordemos (capítulo 2) que Freud llegó a sugerir que las demandas del cuerpo proporcionan el «impulso primario del mecanismo psíquico» y que el prosencéfalo —cuya corteza representa el mundo exterior— es simplemente un «ganglio simpático». En otras palabras, llegó a la conclusión de que las imágenes de la memoria, tanto las conscientes como las inconscientes, se forman «en simpatía» con las demandas del cuerpo y de que solo representamos y aprendemos sobre el mundo exterior perceptivamente porque es en él donde tenemos

que cubrir nuestras necesidades biológicas.

A esta visión del cerebro, la ciencia moderna le ha añadido muchos detalles de los que Freud no tenía la menor pista, y a él lo hemos corregido en varios aspectos importantes. Sin embargo, su conclusión básica es ampliamente aceptada por la neurociencia actual: el cerebro realiza todo un abanico de funciones mentales que no pasan a la conciencia. El título de una famosa revisión de la bibliografía sobre el tema realizada por el científico cognitivo contemporáneo John Kihlstrom lo dice todo: sin duda hay «percepción sin conciencia de lo que se percibe, aprendizaje sin conciencia de lo que se aprende».[128] Otra conocida revisión de John Bargh y Tanya Chartrand resume nuestra comprensión actual con un título más poético: «La insoportable automaticidad del ser»:

La mayor parte de la vida psicológica momento a momento debe de producirse a través de medios no conscientes, si es que se produce. […] Regular de forma consciente y deliberada el comportamiento, las evaluaciones, las decisiones y los estados emocionales propios requiere un esfuerzo considerable y es relativamente lento. Además, parece requerir un recurso limitado que pronto se agota, lo que lleva a pensar que los actos de autorregulación consciente solo pueden producirse de vez en cuando y de forma muy breve. Por otro lado, los [procesos psicológicos…] no conscientes o automáticos no son deliberados, carecen de esfuerzo, son muy rápidos y en muchos casos pueden funcionar en cualquier momento. Lo que es más importante, carecen de esfuerzo y funcionan constantemente para orientar a la persona a fin de que su día transcurra sin complicaciones.[129]

Sin embargo, la ubicuidad del funcionamiento mental inconsciente no fue reconocida a partir de las teorías de Freud, sino por otra vía muy distinta, sobre la base de nuevos hallazgos neurológicos y datos experimentales. Un punto de inflexión fue el caso de H. M., publicado en 1957. Como consecuencia de la resección quirúrgica de su hipocampo —la parte de la corteza responsable de la codificación de los recuerdos declarativos, la parte que le causaba los ataques epilépticos—, era absolutamente incapaz de recordar acontecimientos posteriores a su operación. En cambio, los neuropsicólogos que siguieron su caso desde entonces (1953) hasta su fallecimiento, en 2008, observaron que su rendimiento en las pruebas psicométricas mejoró de forma espectacular con los años, lo que demostraba efectos

claros de la práctica. H. M. aprendió a dominar las pruebas pese a no tener un recuerdo consciente de haberlas hecho.[130]

Con anterioridad, Édouard Claparède ya había hecho observaciones similares, pero no les dio la importancia que merecían. Relató el caso de una mujer amnésica que cada día le estrechaba la mano como si lo acabara de conocer (como mi paciente, el señor S.). Claparède decidió esconderse una chincheta en la mano para ver si la mujer seguía tan dispuesta a saludarlo después de haberse pinchado. En efecto, tras aquella experiencia, la paciente se negó a estrecharle la mano, si bien no tenía ningún recuerdo consciente del doloroso episodio. Al preguntarle por qué se negaba, respondía con evasivas: «¿Acaso una dama no tiene derecho a negar la mano a un caballero?».[131]

Se ha demostrado un procesamiento inconsciente similar para la propia apercepción. Antes de que las juntas de revisión institucional supervisaran la ética de los protocolos de investigación médica, Roger Sperry relató el caso de una mujer a la que, para controlar una epilepsia intratable, le habían separado quirúrgicamente los hemisferios corticales. Cuando le proyectaron imágenes pornográficas en el hemisferio derecho aislado, la mujer se reía y se ruborizaba, aunque el hemisferio izquierdo (verbal) no podía «declarar» de manera consciente a qué se debían esas sensaciones. Solo podía hacer comentarios tangenciales para explicar su rubor, diciendo cosas como «¡Menuda máquina tiene usted ahí, doctor Sperry!».[132]

En este sentido, cabe señalar que el procesamiento mental inconsciente no se limita a las estructuras cerebrales subcorticales. En el caso de Sperry, por ejemplo, la paciente era incapaz de traer a la conciencia la embarazosa información perceptiva que su corteza cerebral derecha había recibido, procesado y reconocido.

Que la corteza puede realizar esas funciones sin conciencia de ello ha quedado demostrado de forma concluyente en muchos estudios experimentales. En uno de ellos, se les enseñaba a los participantes palabras negativas y positivas, pero tan deprisa que no eran conscientes de haber visto nada. Su comportamiento subsiguiente estaba claramente influido por las palabras que decían que no habían visto; por ejemplo, después de que les enseñaran adjetivos negativos en asociación con la fotografía de un rostro A y adjetivos positivos con un rostro B, los participantes mostraban preferencia por el rostro B, aunque no sabían por qué lo preferían.[133] Ese hecho demuestra que tuvieron que ver, leer y comprender de manera inconsciente las palabras negativas y positivas. Teniendo en cuenta que la lectura con comprensión es una función solo cortical —justo del tipo que los

anatomistas clásicos consideraban «mental» por definición—, solo podemos concluir que las funciones corticales no son inherentemente conscientes.

Según la idea más extendida en la ciencia cognitiva actual, la mente interpretada como imágenes de la memoria (ahora llamadas «representaciones») no es intrínsecamente consciente. La parte del cerebro que genera el «contenido» de la conciencia («conciencia como experiencia») —la corteza— puede hacer lo mismo en ausencia de experiencia consciente. Pero, si esto es así, ¿qué hace que sus funciones inconscientes sean mentales? ¿Qué distingue a la corteza —el supuesto órgano de la mente— de otros aparatos de procesamiento de la información, como los teléfonos móviles? Una vez más, estamos al borde de la incoherencia.

Esto nos devuelve a la gran pregunta con la que concluía el capítulo 3: si nuestra conciencia no procede de la corteza, ¿de dónde procede entonces? Es más, si «la mayor parte de nuestra vida psicológica momento a momento» se produce sin experiencia consciente, ¿por qué implica siquiera a la experiencia consciente? ¿Por qué no se produce todo este procesamiento de la información de forma no consciente?

La historia de la neurociencia cognitiva es un cementerio de teorías que intentaron especificar la función de la conciencia. Lo que todas tienen en común es que parten del supuesto de que esa función consiste en «vincular» los numerosos flujos de información dispersos por el cerebro en el todo coherente que caracteriza nuestra experiencia consciente. Por ejemplo, la lectura, el reconocimiento facial, la percepción de los colores, la percepción del movimiento, el reconocimiento de los objetos, la percepción espacial, etc., todos tienen lugar simultáneamente en partes muy dispersas del cerebro. ¿Cómo se aúnan en las imágenes visuales unificadas que percibimos normalmente, con el color y el movimiento de un rostro, por ejemplo, produciéndose en el lugar preciso?

Según Meynert, esta función de vinculación la realizaban las fibras transcorticales, que «asocian» las imágenes de la memoria entre sí. Ha pasado más de un siglo y no hemos avanzado mucho más respecto a esta hipótesis. James Newman y Bernard Baars propusieron que el tálamo genera en la corteza un «espacio de trabajo global» unificado, lo que hace que los diversos fragmentos de información sean globalmente accesibles a la experiencia. Stanislas Dehaene y Lionel Naccache añadieron que las áreas corticales de asociación prefrontal y

parietal integran las actividades de las zonas sensoriales primarias en este espacio de trabajo. Gerald Edelman introdujo los «bucles reentrantes» talamocorticales como la función clave, a través de los cuales se reenvía la información integrada a los niveles anteriores de procesamiento perceptivo. Giulio Tononi hizo hincapié en el procesamiento «masivamente integrado» de la información resultante, afirmando que la clave está en el grado de integración entre los fragmentos; la conciencia es una función de la cantidad de información que se integra. Según la hipótesis de Francis Crick y Christof Koch, la sincronía de las oscilaciones gamma en la corteza vincula y almacena las experiencias; dicho de otro modo: la integración podría tener lugar en el tiempo y no en el espacio. También Rodolfo Llinás sugirió la sincronización de la actividad talamocortical por debajo de cuarenta hercios. Y así sucesivamente. [134]

Ninguna de estas teorías nos explica por qué o cómo la vinculación de la información —ya sea mediante asociación, oscilación, sincronización, reentrada, integración masiva u otros— debería dar lugar necesariamente a la experiencia. ¿Por qué motivo concreto debería experimentarse de forma consciente el contenido de un «espacio de trabajo global» para el procesamiento de la información? ¿Acaso esta vinculación, almacenamiento, sincronización, integración masiva u otros no se producen también en el procesamiento inconsciente de la información? Los ordenadores generan espacios de trabajo globales e integran información de forma masiva todo el tiempo cuando están conectados entre sí a través de internet. Siendo así, ¿por qué no iba a ser también consciente internet?

Una evolución preocupante de estas teorías contemporáneas es que ahora un número cada vez mayor de neurocientíficos respetables (como Koch y Tononi) sugieren que podría serlo. Según ellos, sus teorías los obligan a aceptar esta extraña posibilidad, y al hacerlo se apuntan al giro «panpsiquista» iniciado por Thomas Nagel, quien defiende que todas las cosas podrían ser (solo un poco) conscientes. [135]

Deberíamos evaluar estas teorías a la luz de las observaciones que he comentado antes respecto a la idea de que la conciencia persiste en ausencia de corteza. Newman y Baars ubican su «espacio de trabajo global» en la corteza cerebral, pero hay animales decorticados y niños nacidos sin corteza que parecen ser conscientes. Al menos, las pruebas de que esos animales y esos niños son seres sintientes son mucho más convincentes que las pruebas de que internet también lo es. Además, los adultos con capacidad de habla en los que el sustrato que se

supone que es crucial para el procesamiento de información masivamente integrada ha quedado destruido por la enfermedad nos dicen que su sentido del yo persiste, pese a la ausencia de ese sustrato. Es el caso de mi paciente W., con destrucción total de los lóbulos prefrontales.

Nada de esto debería sorprendernos si combinamos esas observaciones clínicas con los datos experimentales que he comentado antes, según los cuales la corteza realiza casi todo su procesamiento de la información (como la lectura y el reconocimiento facial) de forma inconsciente.

Me gustaría ahora señalar algo importante que queda soterrado en los datos experimentales que he presentado hasta ahora. Cuando la paciente de Claparède se negó a darle la mano, no sabía por qué. Pero es de suponer que sentía aversión al gesto. Seguro que tenía alguna base subjetiva para rechazar la mano del médico, aunque no tuviera acceso a la causa objetiva (el recuerdo de la chincheta). Lo mismo ocurre con la paciente de Sperry, que se sentía avergonzada pero no sabía por qué. Es decir, era inconsciente de la causa objetiva de su sentimiento (las imágenes pornográficas), pero no de las sensaciones subjetivas que las acompañaron (su vergüenza).

Es probable que ocurra lo mismo en el experimento de las palabras: los participantes debieron de sentir cierta preferencia por el rostro B, pero no tenían conciencia del motivo. Lo mismo parece haber sucedido con la niña hidranencefálica de la figura 5, que mostró un placer subjetivo cuando le sentaron a su hermanito en el regazo aunque no era posible que supiera la causa objetiva de su sentimiento, puesto que carecía por completo de imágenes corticales.

Por consiguiente, en todos estos casos, cuando he dicho que los pacientes y los participantes de los estudios que he descrito eran «inconscientes» de las causas de sus actos, he generalizado demasiado. Eran inconscientes de ciertas percepciones o recuerdos —representaciones—, pero sus sentimientos persistían. Cuando emitían sus juicios de valor, seguía habiendo «algo que es como» ser ellos. Eran conscientes de sus sensaciones y sentimientos; solo eran inconscientes del origen de los mismos.

Lo anterior nos da una pista considerable para saber en qué consiste fundamentalmente la sintiencia. El sentimiento, al parecer único entre las demás funciones mentales, es necesariamente consciente. ¿Desde

cuándo es posible que un sentimiento no tenga ninguna cualidad subjetiva? ¿Qué sentido tendría un sentimiento si no lo sentimos? Hasta Freud aceptó que los sentimientos tienen que ser conscientes:

La esencia de una emoción es sin duda que seamos conscientes de ella, esto es, que llegue a conocimiento de la conciencia. En consecuencia, la posibilidad del atributo de inconsciencia quedaría completamente excluida en lo que a las emociones, los sentimientos y los afectos respecta.[136]

Soy consciente de que habrá quienes (incluidos psicoanalistas) no estén de acuerdo con esta afirmación y defiendan la existencia de las emociones inconscientes. Ya volveré luego sobre este punto. Por ahora, permítanme ser muy claro respecto a qué quiero decir con la palabra sentimiento [a veces traducida como «sensación»]: me refiero al aspecto de una emoción (o cualquier afecto) que sentimos. Me refiero al sentimiento, a la sensación. Si es algo que no sentimos, entonces no es un sentimiento.

Unos párrafos antes he dicho que «era de suponer» que la paciente de Claparède sentía cierta aversión, la paciente de Sperry «seguro que» sintió vergüenza y los participantes «debieron de» sentir cierta preferencia. He utilizado estas expresiones porque los neurocientíficos no suelen preguntar por experiencias subjetivas. Y esa es la única razón de que no sepamos qué sintieron los sujetos de los citados estudios. En cambio, yo —como Freud y como Sacks— me tomo muy en serio los informes introspectivos de mis pacientes. De esta forma, espero evitar errores como la confusión entre los sueños y el sueño REM.

Si cien de cada cien personas informan sentir dolor cuando se les pellizca la mano, no es descabellado deducir que pellizcar las manos de la gente causa dolor, aunque tengamos que confiar en informes introspectivos que lo afirmen en cada uno de los casos. Aún más aplicable es esta deducción para las observaciones que podemos replicar en nuestro propio caso, cuando tenemos acceso directo al fenómeno en cuestión. Si pellizcarme la mano me hace sentir dolor, tendré una experiencia personal de lo que quieren decir los demás cuando informan sobre su «dolor».

Los sentimientos son difíciles de investigar porque son inherentemente

subjetivos, pero no por ello podemos permitirnos ignorarlos, como hacen los conductistas. Si excluimos los sentimientos de nuestra explicación del cerebro, nunca entenderemos cómo funciona. El papel que desempeñaban los sentimientos en las fabulaciones del señor S. ilustra muy bien este punto.

Si miramos de cerca lo que aparece en nuestra conciencia, reconoceremos unas cuantas categorías generales de contenido. Hay, por supuesto, «representaciones» del mundo exterior: percepciones, recuerdos y pensamientos sobre ese mundo. Los filósofos han prestado mucha atención a esas representaciones, y el modelo empirista de la mente se diseñó para darles cabida. Sin embargo, no son lo único que encontramos en la conciencia.

También hay sentimientos: sobre lo que sucede en el mundo, sobre lo que pensamos de ese mundo, sobre nosotros mismos (en especial); incluso sentimientos que parecen ser informes acerca del estado de nuestro cuerpo. Asimismo, hay sentimientos que fluyen libres: las emociones y estados de ánimo que caracterizan nuestra experiencia del mundo y moldean nuestro comportamiento en él. A veces se registran como sensaciones corporales, pero hay muchos estados anímicos que no parecen atribuibles ni al estado de nuestro cuerpo ni a nada que podamos identificar en el mundo exterior. ¿No está la conciencia llena de sentimientos como estos? Y, sin embargo, los neurocientíficos que buscan una explicación de la conciencia los han ignorado hasta niveles asombrosos.[137]

En efecto, es increíble hasta qué punto los filósofos empiristas y sus herederos científicos, los conductistas y los científicos cognitivos, ignoraron los sentimientos.[138] Los conductistas afirmaron que todos los aprendizajes se rigen por «recompensas» y «castigos», pero nunca nos contaron qué son esas dos cosas. Realizaron rigurosos experimentos que dieron lugar a la «ley del efecto», según la cual, si un comportamiento va sistemáticamente seguido de recompensas, se incrementará, y si va sistemáticamente seguido de castigos, disminuirá. A este proceso de aprendizaje por experiencia se lo llamó «condicionamiento».

Edward Thorndike, a quien debemos la ley del efecto, quería demostrar que los animales aprenden por ensayo y error, no pensando. Sin embargo, la ley de Thorndike equivale en realidad a una «ley del afecto»,[139] puesto que implica que los comportamientos que nos hacen sentir bien (a nosotros y a otros animales) son los que

repetimos, mientras que los que nos hacen sentir mal son los que evitamos. Así pues, la ley del efecto no es, en su esencia, otra cosa que el «principio de placer» de Freud. Sin embargo, los conductistas no podían aceptar la existencia de algo tan subjetivo como el sentimiento. B. F. Skinner, por ejemplo, declaró de forma notoria que «las “emociones” son perfectos ejemplos de las causas ficticias a las que comúnmente atribuimos el comportamiento».[140]

La redacción original de la ley de Thorndike revela la mentira: «Las respuestas que producen un efecto satisfactorio en una situación particular tienen más probabilidades de repetirse en dicha situación, mientras que las respuestas que producen un efecto desagradable tienen menos probabilidades de repetirse en dicha situación».[141] Más adelante, se sustituyeron las palabras satisfactorio y desagradable por reforzador y castigador. Aquí se explica por qué:

Los nuevos términos, «reforzador» y «castigador», tienen en la psicología un uso distinto al coloquial. Algo que refuerza un comportamiento hace más probable que el mismo se repita, mientras que algo que castiga un comportamiento hace menos probable que ese comportamiento se repita.[142]

Esta definición de los términos es completamente hueca, como no podía ser de otra forma, puesto que ahí la palabra clave es algo. Lo mismo puede decirse de los términos conductistas refuerzo positivo y refuerzo negativo, porque ¿qué hace que un refuerzo sea positivo o negativo, sino un sentimiento?

Se diría que el significado que se le quiere dar a recompensa y castigo es que el valor es inherente al estímulo y no al receptor del estímulo. Si se me acerca un caballo y le doy un terrón de azúcar, es más probable (por la ley del efecto) que se me vuelva a acercar, mientras que si exprimo un limón y le salpica la cara, es menos probable que lo haga. Según Thorndike, el terrón de azúcar y el limón se convierten por sí mismos en recompensas o castigos del comportamiento del caballo; no hay necesidad de considerar las sensaciones o sentimientos que provocan, si es que los hay. El razonamiento es erróneo, claro, porque ubica mal la fuerza causal, lo que lleva al problema que denunciaba Sacks: una mente desprovista de agencia. Una cosa es tratar el cerebro como una caja negra por razones metodológicas, y otra muy distinta es atribuir poderes causales inexistentes a cosas

fuera de la caja y luego concluir que en la caja no pasa nada.

La idea de que los sentimientos son «ficticios» ha supuesto consecuencias nefastas para la ciencia. Por ejemplo, durante la mayor parte del siglo pasado, cuando se investigaban los mecanismos fisiológicos básicos del balance energético, los conductistas prohibieron el uso de las palabras hambre y saciedad por ser conceptos que no se pueden ver ni tocar. Los científicos conductistas solo se permitían hablar de «incentivos» y «recompensas» asociados a la conducta alimentaria. Sin embargo, el tema va más allá de la semántica. Si no se puede utilizar una palabra como hambre, ¿cómo entenderemos el papel que desempeña en la regulación de la alimentación? ¿No podría esto retrasar el desarrollo de tratamientos contra la obesidad centrados en la reducción del hambre? De hecho, la conducta alimentaria está regulada por dos mecanismos cerebrales que interactúan entre sí: un sistema «homeostático», que regula las reservas de energía, y un sistema «hedónico», que media en el apetito.[143] Y si eso es aplicable a efectos corporales como el hambre, también podría aplicarse al desarrollo de tratamientos antidepresivos y ansiolíticos, que podría verse retrasado si se prohibiera el uso de palabras emocionales como tristeza y miedo.[144] Si los sentimientos realmente existen, entonces seguramente tienen correlatos fisiológicos que se pasarán por alto si no los tenemos en cuenta.

A partir de aquí, voy a proceder de forma muy distinta a los conductistas. Siguiendo a Panksepp —el primer neurocientífico que tuvo la temeridad de utilizar palabras como hambre para explicar la regulación del balance energético tanto en humanos como en otros animales—, aceptaré que los sentimientos sí existen.[145] Partiré del supuesto de que ustedes saben en qué consiste sentir sed, tristeza, sueño, diversión, confianza en sí mismos o incertidumbre. Este supuesto no está menos justificado que otras deducciones científicas respecto a elementos de la naturaleza, y se puede someter a prueba de la forma habitual, mediante predicciones falsables.

Los sentimientos son reales y sabemos de ellos porque nos impregnan la conciencia. De hecho, los sentimientos son, por las razones que explicaré a continuación, el manantial del ser sintiente, y en un sentido que apenas me parece metafórico. Desde su origen en algunos de los estratos más antiguos del cerebro, riegan el suelo muerto de las representaciones inconscientes y les confieren vida mental.

[122] Véase Solms y Saling, 1990, para un análisis detallado de los puntos de desacuerdo de Freud con Meynert y otros neuropsicólogos de la época. Dichas discrepancias se recogieron por primera vez en un manuscrito inédito (Freud, 1887), luego aparecieron impresas en Freud, 1888; más tarde las reiteró en Freud, 1891; las siguió desarrollando en Freud, 1893b; y finalmente en su carta a Fliess de 6 de diciembre de 1896.

[123] Freud, 1891, mi traducción. Freud, 1886, p. 14, relató que «visitó varias veces» el laboratorio de Munk en Berlín. En la década de 1880, la propia investigación anatómica de Freud se había centrado bastante en el recorrido de las vías neuronales en el tronco encefálico.

[124] Freud, 1891, mi traducción.

[125] Ibid., mi traducción. Freud también cuestionó la base anatómica y fisiológica de la distinción que Meynert y otros establecieron entre las etapas aperceptiva y asociativa del procesamiento cortical.

[126] Carta a Wilhelm Fliess de 6 de diciembre de 1896 (Freud, 1950a, p. 233).

[127] Los sistemas inconsciente y preconsciente de Freud coinciden por tanto con lo que hoy llamamos memoria a largo plazo «no declarativa» y «declarativa», mientras que su sistema consciente coincide con lo que llamamos «memoria a corto plazo». El primero y el último de los cinco estadios de Freud (percepción y conciencia) son, más que huellas, estados de las neuronas.

[128] Kihlstrom, 1996.

[129] Bargh y Chartrand, 1999, p. 476.

[130] Squire, 2009.

[131] Claparède, 1911.

[132] Galin, 1974, p. 573.

[133] Véase McKeever, 1986.

[134] Crick y Koch, 1990; Newman y Baars, 1993; Dehaene y Naccache, 2001; Bogen, 1995; Edelman, 1990; Marc y Llinás, 1994; Tononi, 2012.

[135] Nagel, 1974. Véanse también Chalmers, 1995a; y Strawson, 2006.

[136] Según Freud, 1915b, p. 177: «Toda la diferencia surge del hecho de que las ideas son catexias —básicamente de huellas mnémicas—, mientras que los afectos y las emociones corresponden a procesos de descarga cuyas manifestaciones finales se perciben como sentimientos. En el estado actual de nuestros conocimientos sobre los afectos y los sentimientos no podemos expresar esta diferencia con mayor claridad».

[137] Por ejemplo, en The Quest for Consciousness, Francis Crick y Christoph Koch explicaron que el afecto implica «los aspectos más difíciles de la conciencia» (Crick, 2004, p. xiv), por lo que «ignoraron de forma deliberada [...] el modo en que las emociones [...] ayudan a formar y moldear la coalición o coaliciones neuronales que son suficientes para la percepción consciente» y se centraron en cambio en «aspectos experimentalmente más manejables de la conciencia», como la percepción visual (Koch, 2004, p. 94). Curiosamente, les pareció más simple el problema del mecanismo neocortical de la visión consciente que el del mecanismo primitivo del afecto en el tronco encefálico, pero reconocieron al mismo tiempo que el afecto da forma a las coaliciones neuronales que son (supuestamente) suficientes para la percepción consciente. En cualquier caso, aun suponiendo que el problema del afecto sea «más difícil», no por ello debería quedar fuera de nuestra explicación de la conciencia. ¡Es como dejar los efectos cuánticos fuera de una explicación de la física!

[138] No quiero decir que lo descuidaran por completo; véase Investigación sobre el conocimiento humano, de Hume (1748).

[139] Término atribuible a Panksepp, 2011.

[140] Skinner, 1953, p. 160.

[141] Thorndike, 1911.

[142] Mazur, 2013.

[143] Encontrarán un fascinante relato de la investigación correspondiente en Leng, 2018, capítulos 16 a 19.

[144] Véase Solms y Panksepp, 2010.

[145] Panksepp, 1974.

05

Sensaciones y sentimientos

Para alegría de sus médicos y de todos nosotros, mi hermano Lee se recuperó bien de la lesión cerebral y se adaptó perfectamente a su nueva vida. Aunque podía ser víctima fácil de los que querían aprovecharse de él, supo defenderse gracias a su tamaño y a su fuerza física. Nunca volvió a ser el mismo, pero siguió adelante con su vida.

A mí, sin embargo, me costó más adaptarme.

Un día mi padre le compró a Lee un reloj de pulsera y le dijo que se lo daría cuando aprendiera a leer la hora. Intentó enseñarle, pero a Lee le costaba entender qué significaba que la manecilla corta estuviera en el número nueve y la larga en el once, por ejemplo. Yo, que tenía entonces cinco años, observaba desde la otra punta de la estancia y adelantaba las respuestas. Mi padre me hizo callar. Aquella noche le dio a Lee el reloj, pese a que no había llegado a entender del todo el funcionamiento. «¿Y yo? ¿Y mi reloj?», pensé.

Al hacerme mayor, aquellos sentimientos infantiles dieron paso a la culpa. A mí me iba bien en la escuela, pero Lee la acabó como pudo. Todo parecía bastante más fácil para mí que para él. Deseaba que mi familia hablara de este problema y me diera estrategias para evitarle disgustos a mi hermano, pero mi madre se negaba rotundamente a hablar sobre el accidente, y mi padre no era el tipo de persona con el que se pudiera hablar de nada que uno sintiera de verdad, y mucho menos de cómo dirigía la familia.

Viéndolo en retrospectiva parece obvio, pero cuando tomé la decisión de estudiar Neuropsicología no se me ocurrió pensar en ningún momento que pudiera tener algo que ver con el accidente de Lee. Sí recuerdo pensar de pequeño que lo único que merecía la pena hacer en la vida era descubrir qué es «ser». Hasta que no recibí mi formación psicoanalítica, años después, cuando ya trabajaba de neuropsicólogo, no até cabos. Era evidente que había elegido una carrera que era un compromiso entre la ambición por un lado y la culpa por el otro. Si me hacía neuropsicólogo, podría tener éxito académico y al mismo tiempo ayudar a las personas con dificultades. Esto también explicaría por qué me resultó tan frustrante la neuropsicología puramente académica y por qué decidí hacerme médico clínico e incluso terapeuta de neurorrehabilitación ocasional, dos papeles profesionales

que tienden a ser menospreciados por los neurocientíficos.

Permítanme que detenga aquí un momento los recuerdos. ¿Les parece creíble este relato de mi decisión? No es ninguna pregunta trampa; creo que mis recuerdos sumergidos de aquellas experiencias explican por qué tomé las decisiones profesionales que tomé. De todas formas, en general, ¿no es así como nuestros sentimientos parecen motivarnos, desde algún punto por debajo del umbral de nuestra conciencia? Es un lugar común de las explicaciones psicológicas, un poco de freudismo que forma ya parte tan intrínseca del sentido común corriente que pocos se atreverían a oponerse a él. Qué extraño resulta entonces afirmar que los sentimientos son conscientes por definición, que en cierta forma son la esencia de la conciencia. ¿Cómo se explica eso?

Un día de 1985, tras acabar mis estudios universitarios, me dirigí al hospital Baragwanath para mi primera jornada como neuropsicólogo residente. El contexto psicosomático previo es que estaba preocupado: mi profesor, Michael Saling, se había ido de permiso sabático a Australia y yo no iba a poder contar con su supervisión habitual. El hecho de perderme de camino al hospital, en Soweto, una ciudad sumamente peligrosa para un chico blanco como yo, no hizo sino agudizar mi estrés.

Cuando por fin encontré el hospital, con el alivio de haber llegado sano y salvo, corrí hasta la Unidad 7 de Neurocirugía y pregunté por el doctor Percy Miller, el especialista neurocirujano al que tenía que presentarme. La enfermera jefe me guio por el pabellón, que desprendía un fuerte olor a una mezcla de fluidos corporales, antiséptico y algo parecido a repollo hervido.

Miller tenía semblante de pájaro. De pie junto a un paciente, mientras extraía líquido cefalorraquídeo de un orificio que le habían practicado en la cabeza, se dirigió a mí efusivamente. «¿Es usted Mark Solms? Encantado. Mike nos dijo que lo sustituiría. Qué bien que haya venido. Un momento, que termino aquí y le enseño la unidad».

Intenté mantener un aire tranquilo y profesional. ¿Sustituirlo? El neurocirujano se lavó las manos y me llevó hasta la primera cabecera de cama. Sin darme tiempo a entender lo que estaba pasando, empezó a relatar el historial clínico del paciente: «Este es el señor Fulano de Tal, cuarenta y tres años, astrocitoma temporal izquierdo, grado 3. Lo operamos el miércoles. Necesito mediciones de referencia de sus funciones de lenguaje. Y supongo que de memoria también; eso lo sabe usted mejor que yo».

Dudé si preguntar o no lo que era un astrocitoma de grado 3.

Pasamos a la siguiente cama. «Este es el señor Mengano, cincuenta y ocho años. Tiene un macroadenoma hipofisario. A este lo tenemos mañana, mejor evalúelo antes que al otro. Haremos un abordaje transesfenoidal. A Mike le parece que los resultados cognitivos son mejores, y tiene sentido».

Y así seguimos, de cama en cama. Cuando íbamos por la quinta o la sexta, ya era demasiado tarde para confesar que no entendía casi nada de lo que estaba diciendo. Tampoco podía recordar los puntos importantes de los primeros casos. En realidad, no recordaba casi nada. Me dije que volvería a mirar los historiales después.

Cruzamos unas puertas batientes y entramos en la unidad de mujeres. El cirujano seguía hablando: «Esta es la señora Fulana de Tal, treinta y seis años; cisticercosis». Tampoco sabía qué era eso. La siguiente paciente estaba completamente desnuda. Algunos de los hombres que había visto también, pero en este caso era distinto. «No debería estar viendo esto», pensé. «¿Saben que no soy médico?». Durante todo el tiempo me estuve fijando en las expresiones faciales de los pacientes: una mirada llorosa, otra en blanco, algunas miradas preocupadas y, en muchos, ninguna mirada.

Todo aquello me superaba. Ellos daban por hecho que yo sabía lo que hacía, y como les seguía la corriente, estaba claro que acabaría siendo responsable de la muerte de alguien.

La unidad pediátrica fue aún peor. El primer niño era un caso de alarmante hidrocefalia; tenía la cabeza como la de E. T., el doble del tamaño normal. Sentí náuseas y calor y empecé a sudar mucho. A la siguiente niña le salía de la nuca un globo de carne fofa. Tenía la mirada fija en la pared. Percy Miller dijo algo sobre un «mielomeningocele». Su voz se alejó, sustituida por los lentos latidos de mi corazón. Se me nubló la vista y, mientras la boca parlante del cirujano se iba borrando, tuve un último pensamiento: «Ahora me golpearé la cabeza contra el suelo, me abriré el cráneo y acabaré en una de estas camas». Sintiéndome extrañamente aliviado, perdí el conocimiento.

Este ejemplo de mi primer día de trabajo ilustra muchas de las características esenciales del «afecto», incluido el hecho de que, aunque no siempre sepamos por qué sentimos lo que sentimos, sí que

sabemos qué sentimos. En mi caso, temí que pasaría lo peor, relacionándolo intelectualmente con mi incompetencia profesional, tras la cual se escondían toda suerte de sentimientos complicados relacionados con mi hermano, como mi necesidad de ayudarlo y, al mismo tiempo, una profunda identificación con él. Sin embargo, la causa inmediata de mi desmayo fue un reflejo corporal primitivo desencadenado directamente por mis emociones.

El síncope vasovagal, o lipotimia, hace que nos desmayemos porque el cerebro reacciona a algo alarmante, por lo general la visión de sangre o algún otro riesgo percibido de lesión física. Este desencadenante (registrado por la amígdala) nos activa el núcleo solitario del tronco encefálico, lo que hace descender de golpe la frecuencia cardíaca y la tensión arterial. Esto provoca a su vez que llegue menos riego sanguíneo al cerebro y nos lleva finalmente a la pérdida de conocimiento.

¿Por qué tenemos este reflejo innato? Pues porque reduce el flujo sanguíneo y, por tanto, detiene la hemorragia, en previsión de una herida.[146] Este reflejo provoca desmayo solo en los seres humanos debido a nuestra postura erguida y al gran tamaño de nuestro cerebro, que requiere más esfuerzo cardíaco (lo que convierte en poco probable la teoría de que los desmayos son una forma de «hacernos los muertos»). En general, el reflejo vasovagal no solo no reduce, sino que aumenta las oportunidades de sobrevivir a las lesiones corporales; y, en cualquier caso, el riego sanguíneo cerebral se restablece en cuanto caemos al suelo.

Veamos otro ejemplo.[147] Normalmente, el control respiratorio es automático; mientras se nos mantengan los niveles de oxígeno y de dióxido de carbono en sangre dentro de unos límites viables, no necesitamos ser conscientes de la respiración para respirar. Sin embargo, cuando los gases en sangre superan esos límites normales, el control respiratorio se inmiscuye en la conciencia en forma de una sensación aguda, el «hambre de aire» o disnea. Los valores inesperados de gases en sangre son una indicación de que es necesario actuar. Es urgente eliminar una obstrucción de las vías respiratorias o salir de una habitación llena de dióxido de carbono. En ese momento, el control respiratorio entra en nuestra conciencia a través de un sistema de alerta interno que experimentamos como alarma, en este caso, concretamente, alarma de asfixia.

Las formas más simples de sentir —hambre, sed, somnolencia, fatiga muscular, náuseas, frío, urgencia urinaria, necesidad de defecar y similares— pueden no parecer afectos, pero eso es lo que son. Lo que

distingue a los estados afectivos de otros estados mentales es que se los dota de una valencia hedónica: sientan «bien» o «mal». En eso se diferencian las sensaciones afectivas, como el hambre y la sed, de las sensoriales, como la vista y el oído. Lo que se ve y lo que se oye no posee valor intrínseco, pero lo que se siente sí.[148]

La bondad o maldad de un sentimiento o sensación nos dice algo sobre el estado de la necesidad biológica que subyace al mismo. Así, tener sed nos hace sentir mal y saciar la sed nos hace sentir bien, porque es necesario para mantener nuestra hidratación dentro de unos límites que sean viables para la supervivencia. Lo mismo ocurre con la sensación desagradable de hambre en relación con el alivio placentero que nos da comer. Resumiendo, el placer y el displacer nos dicen qué tal vamos respecto a nuestras necesidades biológicas. La valencia refleja el sistema de valores en el que se fundamenta toda la vida biológica, esto es, que es «bueno» sobrevivir y reproducirse y «malo» no hacerlo.

Es evidente que lo que motiva a cada cual no son directamente esos valores biológicos, sino las sensaciones subjetivas que generan, aun cuando no tengamos ni idea de cuáles son los valores biológicos subyacentes e incluso no les demos un reconocimiento intelectual. Por ejemplo, comemos cosas dulces porque saben bien, no porque tiendan a tener un contenido energético elevado, que es la razón biológica de que sepan bien. Los afectos nos cuentan largas historias evolutivas que desconocemos por completo. Como en el caso del reflejo vasovagal, solo somos conscientes de las sensaciones.

Si utilizo el término displacer en lugar de dolor es porque en el cerebro hay muchas clases distintas de placer y displacer. El hambre sienta mal y sienta bien aliviarla comiendo; tener el intestino hinchado sienta mal y sienta bien aliviarlo defecando; el dolor sienta mal y sienta bien apartarse de lo que lo causa. Estos son afectos corporales, pero lo mismo es aplicable a los emocionales. La ansiedad por separación sienta mal y respondemos buscando el reencuentro. El miedo sienta mal y lo rehuimos escapando del peligro (y a veces desmayándonos). La alarma por asfixia, el hambre, la somnolencia y el miedo sientan mal por igual, pero de distintas formas. Librarnos de ellos, en cambio, sienta bien, también de distintas formas.

Las distintas sensaciones señalan distintas situaciones con importancia biológica, y cada una de ellas nos lleva a hacer algo distinto: orinar no puede saciar el hambre y comer no puede aliviar una vejiga llena. Recordemos lo que Damasio dijo del paciente B.: «Refería sensaciones de hambre, sed y ganas de evacuar, y también en esos casos se

comportaba en consecuencia». Lo contrario sería fatal.

Los sentimientos y las sensaciones obligan a los seres como nosotros a hacer algo necesario. En ese sentido, son mediciones de demandas de trabajo. El «hambre de aire» (disnea) requiere un trabajo que reequilibre los gases en la sangre. La hipotermia pide un trabajo que devuelva el cuerpo a un rango de temperaturas viable. En la ansiedad por separación, el trabajo necesario ha de conducir al reencuentro con la persona cuidadora. Y así sucesivamente. En la jerga de la teoría de control, los desequilibrios de gases en sangre, la temperatura por debajo de lo normal, la ausencia de los cuidadores o los depredadores que se acercan son «señales de error», y las acciones que provocan están destinadas a corregir los errores. La resolución del afecto mediante algo como la saciedad significa que un error se ha corregido adecuadamente y que a partir de ahí desaparecerá del radar de la conciencia.

Volvemos a transitar un terreno que a los psicoanalistas les resultará extrañamente familiar. Como recordarán, Freud definía la «pulsión» como «medición de la demanda de trabajo que se hace a la mente a consecuencia de su conexión con el cuerpo». Vemos ahora la estrecha relación que hay entre los afectos y las pulsiones: son su manifestación subjetiva. Los afectos son la forma en la que tomamos conciencia de nuestras pulsiones; nos dicen lo bien o lo mal que vamos en relación con las necesidades concretas que miden.

Para eso sirven los afectos, para transmitir qué aspectos biológicos nos están yendo bien o mal e incitarnos a hacer algo al respecto. En este sentido, las sensaciones afectivas son diferentes de las perceptivas. Los filósofos suelen preocuparse por la posibilidad de lo que ellos denominan «inversión de qualia». ¿Cómo sé que el rojo que yo veo es el mismo que ven ustedes? ¿Y si mi rojo lo ven azul? El problema de las otras mentes sugiere que nunca lo sabremos, porque tanto ustedes como yo señalaríamos los mismos objetos del mundo y los llamaríamos «rojo». Pero lo que es cierto para la percepción visual no lo es para la experiencia afectiva. Esa visión del rojo no causa nada diferente que la del azul, por lo que podemos intercambiarlos arbitrariamente sin consecuencias físicas.

No ocurre lo mismo, en cambio, con sensaciones como el hambre en relación con la urgencia urinaria o el miedo en relación con la ansiedad por separación. Una (el miedo) nos empuja a escapar de algo, mientras que otra (la ansiedad por separación) nos empuja a buscar a alguien. La sensación es indisociable del estado corporal que conlleva.[149] Si las intercambiáramos, sentiríamos una urgencia

irresistible por escapar de una persona cuidadora ausente y buscaríamos llorando a un depredador al acecho. Intercambiar la visión subjetiva del rojo con la del azul no tendría consecuencias, pero intercambiar el sentimiento de miedo con la ansiedad por separación (o el hambre con la urgencia urinaria) nos mataría.[150]

Lo segundo que hay que tener en cuenta sobre las sensaciones es que siempre son conscientes. En caso contrario, no son sensaciones. Esto es cierto por definición, como ya he dicho antes, pero también por una característica concreta de la fisiología cerebral. Veremos por qué es así en el próximo capítulo, cuando hablemos de los mecanismos cerebrales de la «excitación». Por ahora, de lo que quiero convencerlos es de que las sensaciones siempre son conscientes, sin excepción. No quiero decir con ello que todos los mecanismos cerebrales que regulan necesidades sean conscientes, pero sí que la diferencia radica en si la necesidad se siente o no. Puede que nuestra proporción de agua y sal esté bajando todo el tiempo, en segundo plano, pero cuando lo sentimos, es cuando queremos beber. Puede que estemos objetivamente en peligro sin saberlo, pero cuando lo sentimos, es cuando buscamos la forma de escapar.

Las cosas distintas necesitan nombres distintos, y la diferencia entre necesidades sentidas y no sentidas hace necesario introducir una distinción terminológica: las «necesidades» no son lo mismo que los «afectos». Las necesidades corporales pueden registrarse y regularse de forma autónoma, como en los ejemplos del control cardiovascular y respiratorio, la termorregulación y el metabolismo de la glucosa. A estas funciones se las llama «vegetativas», y con propiedad, porque en ellas no hay nada consciente. De ahí el término reflejo autónomo. La conciencia solamente entra en la ecuación cuando las necesidades se sienten, que es cuando nos formulan demandas de trabajo. (Nota: las pulsiones miden las demandas de trabajo que se hacen a la mente, mientras que algunas necesidades nunca activan la acción voluntaria y, por lo tanto, nunca se vuelven conscientes. La tensión arterial es un ejemplo nada desdeñable clínicamente: no sabemos si ha bajado o subido en exceso hasta que ya es demasiado tarde).

Por otra parte, las necesidades emocionales también pueden gestionarse automáticamente, mediante estereotipos de comportamiento como los «instintos» (estrategias innatas de supervivencia y reproducción que Freud puso en el centro de su concepción de la mente inconsciente). Sin embargo, por los motivos que expondré más adelante en este mismo capítulo, las necesidades emocionales suelen ser más difíciles de satisfacer que las corporales. Por ello las sensaciones que provocan suelen prolongarse más en el

tiempo. Una sensación desaparece de la conciencia cuando la necesidad que anuncia se ha visto satisfecha.

Lo tercero que cabe señalar sobre las sensaciones es que las necesidades sentidas reciben un trato prioritario frente a las no sentidas. Nos acucian constantemente necesidades de todo tipo. Las funciones vegetativas, como el balance energético, el control respiratorio, la digestión, la termorregulación, etc., están siempre activas, como también lo están los comportamientos estereotipados de diversa índole. Y todo eso no podemos sentirlo a la vez, entre otras razones porque solo podemos hacer una cosa (o unas pocas cosas) cada vez. Hay que elegir. Y la selección se hace a partir del contexto. La prioridad la determina la intensidad relativa de cada necesidad (el tamaño de las señales de error) en relación con el abanico de oportunidades que nos permiten las circunstancias de cada momento. Veamos un ejemplo sencillo: mientras doy una conferencia, se me va llenando la vejiga, pero no noto la creciente presión hasta que termina la conferencia, y en ese instante siento repentinas ganas de orinar. Me hago consciente de la señal de error por el cambio contextual. En ese momento, mi necesidad corporal presiente la oportunidad y salta a la conciencia en forma de sensación.[151]

El hecho de dar prioridad a una necesidad frente a otra tiene consecuencias notables, siendo esta la más importante: cuando nos hacemos conscientes de una necesidad, cuando la sentimos, rige nuestro comportamiento voluntario.

¿Qué significa «voluntario»? Lo contrario de «automático». Significa estar sujetos a decisiones de aquí y ahora. Y las decisiones tienen que estar basadas en un sistema de valores, lo que determina la «bondad» frente a la «maldad» de las cosas. De lo contrario, nuestras respuestas a los sucesos desconocidos serían aleatorias. Cerramos así el círculo, porque hemos vuelto a la característica más fundamental del afecto: su valencia. Decidimos lo que hacer y lo que no hacer en base a las consecuencias sentidas de nuestros actos. Es la ley del afecto. El comportamiento voluntario, guiado por el afecto, goza así de una ventaja adaptativa enorme sobre el comportamiento involuntario; nos libera de los grilletes de la automaticidad y nos permite sobrevivir en situaciones imprevistas.

Que el comportamiento voluntario tenga que ser consciente revela la función biológica más profunda del sentimiento o sensación, que consiste en guiar nuestro comportamiento en condiciones de incertidumbre. Nos permite determinar en caliente si es mejor o peor emprender una acción que otra. En el ejemplo de la disnea, la

regulación de los gases en sangre se vuelve consciente cuando no tenemos una solución a punto para mantener nuestros límites fisiológicamente viables. En nuestra prisa por escapar de la habitación llena de dióxido de carbono, por ejemplo, ¿cómo saber hacia dónde ir? No hemos estado nunca en esa situación (en ningún edificio en llamas, ni mucho menos en este), lo que hace imposible predecir qué hay que hacer. Ahora tenemos que decidir si vamos por aquí o por allá, hacia arriba o hacia abajo, *etc.* Tomamos esas decisiones recorriendo el problema con nuestras sensaciones; por ejemplo, la sensación de asfixia aumenta o disminuye en función de si hemos tomado el camino correcto o no, esto es, según aumenta o disminuye la disponibilidad de oxígeno.

La sensación consciente de alarma por asfixia implica estructuras neuronales y sustancias químicas distintas de las responsables del control respiratorio inconsciente, de la misma forma que la sensación de hambre activa sistemas cerebrales distintos de los responsables de la regulación autónoma del balance energético. La ciencia nunca lo habría descubierto si hubiera seguido ignorando las sensaciones.

Podríamos hablar mucho más sobre las sensaciones; por ejemplo, sobre la manera en que permiten aprender de la experiencia (mediante la ley del afecto) y se relacionan con el pensamiento, pero los puntos que he desgranado brevemente ya explican por qué sentimos. He descrito lo que las sensaciones añaden al repertorio de mecanismos que los seres vivos utilizamos para sobrevivir y reproducirnos. Es la contribución de la psicología a la biología. La selección natural determinó estos mecanismos de supervivencia, pero después de que evolucionaran los sentimientos —a saber, la capacidad única que tenemos como organismos complejos de registrar nuestros propios estados—, apareció algo completamente nuevo en el universo: el ser subjetivo.

Cuesta imaginar cómo se puede tener dolor físico sin sentirlo; tener dolor es, justamente, sentirlo. Pero ¿y los estados afectivos más sutiles que llamamos «emociones» y «estados de ánimo»? ¿Hace falta sentir que se es feliz para serlo? ¿Cuántas veces hemos descubierto tarde que estábamos de mal humor, o que tal vez teníamos una depresión?

Hemos empezado a explorar el afecto a través de sus formas corporales, pero es porque nos dan los ejemplos más sencillos y porque sin duda fueron también las primeras en aparecer en la evolución. Creo que el «amanecer de la conciencia» apenas fue algo

más que sensaciones somáticas valenciadas. Lo que quiero mostrar ahora es que las emociones humanas son versiones complejas de algo parecido. En última instancia, también son «señales de error» que registran desviaciones de nuestros estados biológicamente preferidos, que nos dicen si las medidas que tomamos están mejorando o empeorando las cosas.

Lamentablemente, en la neuropsicología actual no hay ninguna clasificación de los afectos que cuente con el acuerdo de todos. Yo acabo de distinguir entre afectos corporales y afectos emocionales, pero en la naturaleza no existen delimitaciones tan radicales. Lo que hago con esa distinción es seguir la taxonomía de Jaak Panksepp, que goza de aceptación, si no universal, al menos general. Panksepp clasificó después la gran variedad de afectos corporales en subtipos interoceptivos («homeostáticos») y exteroceptivos («sensoriales»). El hambre y la sed, por ejemplo, son afectos homeostáticos, mientras que el dolor, la sorpresa y el asco son afectos sensoriales. (No utilizo el término homeostático de Panksepp para el subtipo interoceptivo del afecto corporal porque, como veremos enseguida, todos los afectos son homeostáticos, y la acepción más restringida de Panksepp resultaría aquí confusa). Resumiéndolo, según él hay tres tipos de afecto: homeostáticos, sensoriales (ambos corporales) y emocionales (que implican el cuerpo pero que no pueden describirse como «corporales» en un sentido simple). Imaginemos, por ejemplo, que echamos de menos a un hermano, que es un estado emocional: no es corporal como lo son el hambre y el dolor.

Panksepp basó su taxonomía en estudios de estimulación cerebral profunda que realizó junto a sus alumnos con miles de animales, como hicieron muchos otros antes. Visité su laboratorio muchas veces y era un auténtico zoológico, lleno unos días de palomas y pollos, y otros de perros beagle, cobayas, ratas o topillos de las praderas. (La enfermedad de la que finalmente murió podría haber sido causada por un exceso de exposición a algunos de aquellos animales, sobre todo las aves).

No hay la menor duda de que Jaak sentía compasión, incluso amor, por esos animales, pero tampoco podemos negar que fueron sacrificados a la ciencia en gran número sin haber elegido ellos ese destino. Es una triste ironía que debamos a la investigación de Panksepp el conocimiento casi seguro de que los animales antes mencionados son seres sensibles, sujetos a emociones intensas, que en esencia no son distintas a las que sentimos nosotros. Como resultado de esos hallazgos y de la creciente preocupación por los aspectos éticos que de ellos se derivan, Panksepp pasó las últimas décadas de

su vida estudiando únicamente emociones positivas.

De aquí en adelante alternaré observaciones sobre animales con observaciones sobre humanos, y será deliberado. Como Panksepp dijo cuando los colegas lo acusaron de antropomorfismo con los animales: antes preferiría declararse culpable de zoomorfismo con los humanos. La finalidad de sus experimentos era determinar qué estructuras y circuitos cerebrales despertaban de forma fiable las mismas respuestas afectivas, no solo en distintos individuos, sino también en distintas especies. En el caso de los afectos emocionales, se vio que siete de ellos se podían reproducir fiablemente no solo en todos los primates, sino también en todos los mamíferos, mediante la estimulación de exactamente las mismas estructuras y sustancias químicas cerebrales. (Muchos de ellos también pueden provocarse en las aves y algunos en todos los vertebrados). Los mamíferos se separaron de las aves hace unos doscientos millones de años; luego estas emociones tienen la misma antigüedad. En cualquier caso, como los humanos son mamíferos, yo me centraré en esos siete tipos. Por lo que sabemos, son los ingredientes básicos de todo el repertorio emocional humano. Todo nuestro gran baúl de penas y alegrías parece ser el resultado de estos siete sistemas, que se mezclan entre sí y con procesos cognitivos superiores.

Posiblemente, la taxonomía alternativa más conocida de las «emociones básicas» (como se las denomina) sea la de Paul Ekman.[152] Las disparidades se deben sobre todo a que Ekman utilizó un método diferente al de Panksepp, a saber, el estudio de las expresiones faciales y los comportamientos relacionados. Como ya había constatado Charles Darwin en La expresión de las emociones en el hombre y en los animales (1872), las distintas especies de mamíferos comparten muchas expresiones notables. Sin embargo, pese a que Ekman y Panksepp no clasifican igual los afectos —por ejemplo, para el primero el asco es un afecto «emocional», mientras que para el segundo es «sensorial»—, existe un amplio acuerdo sobre los afectos en sí. Nadie cuestiona seriamente que el asco exista, así que en cierto modo no importa demasiado cómo lo clasifiquemos.

La principal voz discordante es la de Lisa Feldman Barrett. También aquí el desacuerdo es sobre todo atribuible a diferencias metodológicas.[153] Barrett se centra en emociones autoinformadas en humanos y descubre, como es de esperar, que hay una enorme variabilidad en la caracterización y análisis de los sentimientos según las personas (y las culturas). Eso no es óbice para que debajo de la superficie socialmente construida acechen los tipos básicos naturales. Enseguida ilustraré los mecanismos por los que surge esa variabilidad,

pero, de forma resumida, podemos decir que nuestros reflejos e instintos proporcionan herramientas básicas para la supervivencia y el éxito reproductivo aun cuando no pueden equiparnos bien para la miríada de situaciones y entornos imprevistos en los que nos podemos encontrar. Necesitamos por tanto complementar adaptativamente las respuestas innatas mediante el aprendizaje por experiencia. El hecho de que los seres humanos lo hagamos con tanta facilidad es el principal motivo de que, no siempre para bien, hayamos acabado dominando el mundo en el grado en el que lo hacemos ahora.

Los programas instintivos en los que se apoyan las acciones de los humanos suelen estar tan condicionados por el aprendizaje que ya no son reconocibles como «instintivos». Sin embargo, los instintos y los reflejos siguen estando siempre en el fondo de la escena. Toda la teoría psicoanalítica se basa en esta idea: si nos molestamos en buscarlas, siempre podremos distinguir tendencias instintivas implícitas detrás de las intenciones explícitas.

Ahora ofreceré una breve introducción a las emociones instintivas según la taxonomía de Panksepp.[154] Si en la forma de clasificar los afectos hay diferencias entre los científicos, también las hay en cómo los llaman. Panksepp escribe el nombre de las emociones básicas en mayúsculas para distinguirlas del uso coloquial, es decir, para indicar que se refiere a funciones cerebrales completas, no solo a los sentimientos.

(1) LUJURIA. No estamos sexualmente excitados todo el tiempo. El sentimiento erótico entra en la conciencia solo cuando damos prioridad al sexo sobre otras motivaciones, lo cual ocurre en el contexto de necesidades y oportunidades variables. Cuando se despierta el deseo sexual, lo sentimos y actuamos guiados por los sentimientos eróticos. No prestamos atención a los mismos detalles cuando estamos sexualmente excitados que cuando tenemos miedo, por ejemplo, ni tampoco actuamos igual. Así, nuestra conciencia exteroceptiva y nuestro comportamiento voluntario están determinados por nuestro estado interior; experimentamos el mundo de forma distinta —trayendo a la mente experiencias distintas— en función de lo que sentimos. Esto explica lo difícil que es sentir al mismo tiempo atracción sexual y repulsión por miedo; no podemos dar prioridad a ambas cosas. Cuando la prioridad la tiene la necesidad de ponernos a salvo, las motivaciones sexuales se borran de la mente.

No está claro si la LUJURIA debe clasificarse como afecto «corporal» o «emocional». Hay quien incluso cuestiona que la sexualidad sea una necesidad. Estamos ante un ejemplo excelente de la diferencia entre unas necesidades (inconscientes) y los afectos (conscientes) que provocan. Cuando realizamos actos sexuales, por lo general no estamos intentando cumplir nuestro deber biológico. De hecho, la mayoría de las veces, confiamos en no reproducirnos. Del mismo modo que la apetencia de algo dulce no tiene por qué significar una búsqueda de suministro de energía, lo que nos motiva como seres subjetivos es buscar el placer erótico, no el éxito reproductivo. Es decir, lo que nos impulsa son sensaciones y sentimientos. Sin embargo, los organismos vivos necesitan reproducirse, al menos por término medio. Es el principal motivo de que el sexo se volviera subjetivamente placentero gracias a la selección natural.

He dicho «por término medio» porque no todas las actividades sexuales llevan a la reproducción, sino solo las suficientes para conservar la especie; este dato ejemplifica otro principio central: la utilidad limitada de las conductas innatas a la hora de cubrir nuestras necesidades emocionales. En el sexo, los aspectos innatos se reducen a poco más que la turgencia y la lubricación genitales, la lordosis (el arqueo de la espalda, que ofrece la vagina a la penetración), la monta, la penetración, el empuje y la eyaculación. Junto a estos reflejos, acariciar el clítoris o el pene (que son órganos equivalentes) a un ritmo determinado produce sensaciones placenteras que anuncian la liberación de la tensión sexual, por medio del orgasmo, hasta la saciedad. Estos artificios involuntarios no nos equipan para la difícil tarea de persuadir a otras personas —sobre todo aquellas por las que sentimos atracción— para que satisfagan nuestro deseo de tener relaciones sexuales con ellas. La razón principal de que sea más difícil cubrir las necesidades emocionales que las corporales es que suelen implicar a otros agentes sensibles, que tienen sus propias necesidades; no son meras sustancias como la comida o el agua. Por lo tanto, para satisfacer las necesidades sexuales tenemos que complementar nuestro conocimiento innato con otras habilidades, adquiridas a través del aprendizaje. Solo eso ya explica la amplia variedad de actividades sexuales que practicamos, además de la forma «corriente» que se nos legó por selección natural.

Cabe observar que el aprendizaje no borra los reflejos y los instintos; los desarrolla, complementa y anula, pero siguen ahí. Las farolas de la calle iluminan por la noche, pero no por eso suprimen la oscuridad. El mecanismo habitual para actualizar los recuerdos a largo plazo, la «reconsolidación» —que describiré en el capítulo 10—, no es aplicable a los reflejos y los instintos por la sencilla razón de que los reflejos y

los instintos no son recuerdos, son inclinaciones fundamentales que se han «inculcado» en cada especie a través de la selección natural.

Nuestra gama de conductas sexuales se amplía aún más por el hecho de que los circuitos cerebrales tanto para la LUJURIA típica del sexo femenino como para la típica del sexo masculino existen en todos los mamíferos. La tendencia que acaba imponiéndose la determinan diversos factores, como los sucesos genéticos e intrauterinos.[155] Ahora no entraré en los detalles anatómicos y químicos, pero sí señalaré que ambos circuitos surgen en el hipotálamo y terminan en la sustancia gris periacueductal o SGP. (En la figura 6 se muestra la ubicación de la SGP, que enseguida veremos por qué es tan importante). Es decir, estos circuitos son completamente subcorticales.[156]

(2) BÚSQUEDA. Todas las necesidades corporales (y sexuales), que son registradas por «detectores de necesidades» situados principalmente en el hipotálamo medial, activan esta segunda pulsión emocional, que hemos visto en el capítulo 1. Es casi sinónimo del concepto freudiano de «libido», pero Freud no sabía que la LUJURIA simplemente activa este sistema; la LUJURIA y la BÚSQUEDA no son lo mismo. La BÚSQUEDA genera un comportamiento de «búsqueda de alimento», acompañada por un estado de sentimiento consciente que se puede caracterizar como expectación, interés, curiosidad, entusiasmo u optimismo. Imaginemos un perro en un campo abierto; con independencia de cuáles sean sus necesidades corporales en ese momento, la búsqueda de alimento lo empuja a relacionarse de forma positiva con el entorno para satisfacerlas allí. Casi todo lo que necesitamos los seres vivos está «ahí fuera»; mediante la búsqueda de alimento aprendemos, casi por accidente, qué cosas del mundo satisfacen cada una de nuestras necesidades. Así, codificamos sus relaciones causa-efecto, un hecho que vuelve a ilustrar cómo los instintos estereotipados conducen a un aprendizaje individualizado.

Lo que distingue a la BÚSQUEDA de las demás emociones básicas es que actúa de forma proactiva respecto a la incertidumbre. De ahí surgen los comportamientos que nos hacen buscar lo nuevo e incluso correr riesgos. La búsqueda de alimento nos hace explorar cosas interesantes para luego saber a qué atenernos cuando volvamos a encontrarlas. Cuando un perro ha explorado un seto, por ejemplo, y se ha familiarizado con su contenido, la próxima vez sentirá menos interés por él. En consecuencia, la BÚSQUEDA es nuestra emoción «por defecto». Cuando no estamos atrapados por alguno de los demás

afectos («relacionados con tareas»), nuestra conciencia tiende hacia esta sensación generalizada de interés por el mundo.

En términos anatómicos, las neuronas del circuito de BÚSQUEDA proceden del área tegmental ventral del tronco encefálico, desde donde ascienden por el hipotálamo lateral hasta el núcleo accumbens, la amígdala y la corteza frontal (véase fig. 2). En términos químicos, su neuromodulador determinante es la dopamina, la «materia de los sueños» (véase el capítulo 1).[157] Esto revela un dato interesante sobre la BÚSQUEDA, a saber, que puede despertarse incluso durante el sueño por demandas de trabajo que se hacen a la mente, lo que conduce a actividades de resolución de problemas que deben guiarse por sentimientos conscientes. De ahí que soñemos.

(3) IRA. Mientras nos relacionamos positivamente con el mundo a través de la BÚSQUEDA, en la creencia optimista de que así quedarán cubiertas nuestras necesidades, no siempre nos sale todo bien. Al igual que la prehistoria evolutiva nos dotó de reflejos e instintos que predicen de forma fiable cómo satisfacer nuestras necesidades corporales, también nacemos con tendencias emocionales que predicen cómo sacarnos de apuros. En situaciones difíciles de importancia universal, damos prioridad a los sentimientos pertinentes por los que se regirá el comportamiento. Eso nos ahorra el coste biológico de tener que reinventar las ruedas que permitieron a nuestros antepasados sobrevivir y reproducirse. Las emociones son un valioso legado: transmiten técnicas de supervivencia innatas —conocimiento implícito e inconsciente— en forma consciente de sentimientos que pueden guiar de manera explícita nuestras acciones.

Cuando se desencadena el sistema de la IRA —por cualquier cosa que se interponga entre nosotros y lo que sea que en esos momentos cubriría nuestras necesidades—, nuestra conciencia se ve invadida por sentimientos que van desde la frustración irritada hasta la furia desatada. Los reflejos e instintos que se liberan en esos casos son, entre otros, la piloerección (los pelos de punta), la protrusión de las uñas, los siseos, los gruñidos y enseñar los dientes, seguido del «ataque afectivo», consistente en abalanzarnos sobre el blanco de nuestra ira y morderlo, pegarle y darle patadas hasta que ceda.

¿Por qué sentimos los afectos que acompañan semejante comportamiento? La respuesta se repite: los sentimientos nos dicen cómo nos va, si las cosas están saliendo bien o mal mientras intentamos librarnos de un obstáculo que, en muchos casos, también

se está intentando librar de nosotros. Sentimos el dulce sabor de la victoria o la amargura de la derrota, lo que nos guiará respecto a qué hacer después, siendo posible que el dolor (un afecto sensorial que se suprime durante el ataque afectivo) reclame un lugar prioritario y eso nos haga sustituir la IRA para poner fin a la lucha y tal vez iniciar la huida.

¿Cómo puede funcionar todo esto automáticamente, sin una evaluación consciente constante? La pregunta es aún más aplicable al rol del afecto en el pensamiento, un tema que podemos aprovechar para introducir en este momento. El pensamiento es una acción «virtual»; la capacidad de probar cosas con la imaginación; una capacidad que, por razones biológicas obvias, salva vidas. Esta capacidad no es exclusiva de los humanos, pero en nosotros está especialmente desarrollada. Veamos, pues, un caso humano. Imaginemos la siguiente situación, extraída de mi experiencia. El director de mi escuela me está dando una reprimenda en su despacho. Yo me siento cada vez más enfadado. La respuesta instintiva es el ataque afectivo, pero pienso en las consecuencias potenciales. En lugar de abalanzarme sobre él, inhibo mi tendencia a la acción instintiva e imagino mi abanico de alternativas y las exploro sintiéndolas. Al final, llego a una solución satisfactoria; tras salir de su despacho, cuando nadie mira, le pincho las ruedas del coche. De esa forma, reduzco mi IRA sin sufrir duras consecuencias. El ejemplo vuelve a ilustrar por qué los estereotipos de comportamiento innatos deben complementarse con el aprendizaje por experiencia, incluida la forma imaginaria de la experiencia llamada «pensamiento». Cuando nos enfrentamos a frustraciones de la vida real, que suelen contener necesidades en conflicto (en este caso, IRA frente a MIEDO), las soluciones instintivas no nos bastan. Pero insisto: complementar las respuestas instintivas con el aprendizaje no las suprime. Yo decidí no atacar a mi director, pero las ganas de hacerlo siguieron ahí y volverían a surgir en situaciones similares en el futuro. (El ejemplo no sirve solo para seres humanos. Un perro no llegaría a una solución como esa, pero los primates inventan todo tipo de astutas estrategias). [158]

Emociones como la IRA no son «meros» sentimientos. Las emociones desempeñan un rol fundamental en la supervivencia. Imaginemos las consecuencias si no pudiéramos reclamar los recursos disponibles ni impedir que otros se llevaran nuestra parte. Si no pudiéramos frustrarnos, irritarnos o enfadarnos, no sentiríamos inclinación a luchar por lo que necesitamos, en cuyo caso, tarde o temprano estaríamos muertos. Es fácil ignorar la función biológica de la emoción en las condiciones civilizadas en las que vivimos hoy día, pero solo

llevamos unos doce mil años viviendo así (en asentamientos permanentes con leyes artificiales que regulan el comportamiento social). La civilización, una característica muy reciente en nuestra existencia como mamíferos, no tuvo ningún papel en el diseño de nuestro cerebro.

El pensamiento consciente requiere la corteza cerebral, pero los sentimientos que lo guían no. El circuito que media en la IRA es subcortical casi en su totalidad, y, al igual que los demás circuitos afectivos, su destino final es la SGP del tronco encefálico.[159]

4) MIEDO. La dicotomía lucha-huida demuestra que el ataque afectivo no siempre es la mejor forma de enfrentarse al adversario. Los factores contextuales que separan la lucha de la huida están codificados en la amígdala, que media tanto para la IRA como para el MIEDO.[160]

La mayoría de los mamíferos «saben» desde el primer día que hay cosas que producen un miedo intrínseco. Los roedores recién nacidos, por ejemplo, se paralizan de miedo ante un simple pelo de gato, aunque no tengan ninguna experiencia con los gatos y desconozcan su actitud hacia los ratones. El motivo es fácil de entender: si cada ratón tuviera que aprender por experiencia cómo reaccionar ante un gato, sería el fin de los ratones. Una vez más, vemos el enorme valor biológico de las emociones.

Los humanos tememos peligros como las alturas, la oscuridad y los animales que culebrean y se arrastran hacia nosotros, y los evitamos gracias a los mismos instintos y reflejos que otros mamíferos, con comportamientos de parálisis y huida. Contrariamente al reflejo vasovagal de interrupción que hemos visto antes, estos comportamientos de «escape» se ven facilitados por la respiración acelerada, el aumento de la frecuencia cardíaca y el redireccionamiento de la sangre desde los intestinos hacia la musculatura esquelética. (Lo que explica la pérdida del control de esfínteres asociada con el miedo intenso). Como ocurre con otras emociones, el sentimiento consciente del miedo nos indica si estamos cerca o lejos de ponernos a salvo.

Encontramos un ejemplo interesante en el caso de la paciente S. M., que se publicó por primera vez cuando ella se acercaba a la treintena. Padecía la enfermedad de Urbach-Wiethe, un trastorno genético raro que provoca la calcificación de la amígdala. S. M. no sentía miedo. En consecuencia:

[S. M.] ha sido víctima de numerosos actos delictivos y encuentros traumáticos que han puesto en peligro su vida. La han atracado a punta de navaja y de pistola, casi la matan en un incidente de violencia doméstica y ha recibido amenazas de muerte explícitas en numerosas ocasiones. Pese al peligro que corría su vida en muchas de estas situaciones, S. M. nunca mostró señales de desesperación, urgencia u otras respuestas conductuales normalmente asociadas a incidentes semejantes. El número desproporcionado de sucesos traumáticos en la vida de S. M. se ha atribuido a […] una marcada incapacidad por su parte para detectar amenazas inminentes en su entorno y aprender a alejarse de situaciones potencialmente peligrosas.[161]

Yo he estudiado a un gran número de pacientes similares, porque cerca de donde nací, en un rincón remoto de Sudáfrica llamado Namaqualand, se registra una incidencia excepcional de la enfermedad de Urbach-Wiethe. (El gen defectuoso llegó allí con un colono alemán y se quedó concentrado en una comunidad aislada). Me resultaron especialmente interesantes sus sueños, porque son breves, simples y expresan deseos. Una de las pacientes que estudié, cuyo marido estaba en paro, soñó: «Mi marido encontraba trabajo y yo me sentía muy feliz». Otra tenía una hija con discapacidad y soñó: «Mi hija podía andar y yo me sentía muy feliz». Otra, que había perdido a su padre, soñó: «Volví a ver a mi padre y me sentía muy feliz». Son sueños típicos de pacientes de Urbach-Wiethe, cuya imaginación carece de miedo y no prevé peligro alguno en el cumplimiento de sus deseos.[162]

En principio, la mayoría de nosotros hemos nacido con desencadenantes específicos del MIEDO. Imaginen, si no, qué ocurriría si todos tuviéramos que aprender por experiencia lo que ocurre si saltamos de un acantilado o si atrapamos una víbora con la mano. Por eso descendemos de antepasados que no sintieron ninguna inclinación por intentarlo. Los que lo hicieron no son nuestros antepasados porque no dejaron descendencia alguna. Tenemos muchos motivos para agradecer esa herencia.

Con todo, debemos aprender a qué más hay que temer. Aprendemos por experiencia —incluido el pensamiento— que hay otras cosas aparte de las alturas de vértigo y las mordeduras de serpiente que nos pueden hacer mucho daño. Los enchufes eléctricos y las descargas que

producen, por ejemplo, no podrían haberse previsto evolutivamente y, en cambio, son tan peligrosos como las víboras. Para complementar nuestras respuestas instintivas, tenemos que aprender qué más hay que hacer cuando sentimos miedo. Quedarnos paralizados o salir corriendo de todo lo que nos asusta no es un comportamiento de adaptación, como tampoco lo es atacar a todo lo que nos frustra. A estas alturas debería estar claro qué papel tienen los sentimientos conscientes en este proceso de aprendizaje: nos dicen lo que funciona y lo que no antes de que sea demasiado tarde, y así nos ayudan a seguir con vida.

El condicionamiento por miedo revela otros datos importantes sobre qué es consciente y qué no. Una de sus características especiales es el «aprendizaje por exposición única». Basta con meter el dedo en un enchufe una sola vez para que no lo hagamos nunca más. Y es fácil entender por qué: si hemos tenido la suerte de sobrevivir la primera vez, ¿para qué repetir la experiencia? Sin embargo, como ocurre con todos los demás mecanismos biológicos que sustentan las emociones, no es necesario saberlo en ningún sentido «declarativo»; el condicionamiento se produce de forma automática. La razón es que el condicionamiento por MIEDO no requiere la participación de la corteza cerebral. Puede producirse en la primera infancia, incluso antes de que madure el hipocampo (la estructura cortical responsable de depositar los recuerdos a largo plazo declarativos). Esta es la causa de que, como en el caso de Claparède, tantas personas neurológicamente sanas teman cosas sin saber por qué.

Los científicos cognitivos atribuyen lo anterior al aprendizaje «inconsciente», pero, como ya hemos visto, solo porque pasan por alto el afecto. Si bien es cierto que muchas personas son inconscientes de los motivos por los que tienen miedo a algunas cosas, también lo es que son muy conscientes de los sentimientos asociados. Y los sentimientos son lo único que hace falta para guiar el comportamiento voluntario. Volvamos al experimento de las palabras que habíamos visto: si asociamos de manera subliminal palabras como asesino y violador con el rostro A, y cariñoso y generoso con el rostro B, los sujetos del estudio sentirán una preferencia cuando se les pida que elijan uno u otro, aunque no sepan por qué. Será una decisión guiada por una corazonada, pero, como los sentimientos suelen pasarse por alto, se describirá cognitivamente con palabras como suposición.[163]

Esto explica bastante bien la perplejidad que despiertan las «emociones inconscientes» en la ciencia cognitiva. No son tanto las emociones las que son inconscientes, sino los elementos cognitivos que las despiertan. Como ya hemos visto cuando hablábamos del

pensamiento, saber a qué se deben nuestros sentimientos puede ser muy útil, pero no es esencial. De hecho, a veces es mejor no pensar antes de actuar, entre otras razones porque pensar lleva tiempo.

En el condicionamiento por miedo pasa lo mismo. Cuando hemos aprendido a temer algo —sobre todo si no sabemos conscientemente por qué—, la asociación es casi irreversible. Como dijo de forma memorable Joseph LeDoux, los recuerdos de miedo son «indelebles». [164] Más adelante trataré los importantes aspectos que ese hecho revela sobre la memoria inconsciente en general, pero por ahora baste con decir que los recuerdos «no declarativos» (como los emocionales y los procedimentales) son difíciles de olvidar, por la misma razón por la que son inconscientes: suponen menos incertidumbre —es decir, son más generalizables—, por lo que están menos sujetos a la revisión contextual. Así es como los comportamientos adquiridos se vuelven estereotipados y automatizados. En tanto en cuanto la finalidad de la cognición es aprender a satisfacer nuestras necesidades en el mundo, la automatización es el ideal del aprendizaje.

(5) PÁNICO-DOLOR. La ansiedad por separación es distinta del MIEDO. Surge por primera vez en la infancia, cuando nos vinculamos instintivamente a la madre (o a la principal persona cuidadora). A diferencia del condicionamiento por miedo, pero por motivos biológicos igual de válidos, tarda unos seis meses en desarrollarse; un solo episodio de crianza no basta para demostrar que se puede confiar en alguien para siempre.

Cuando se separa a los mamíferos de sus figuras de apego, se reproduce una secuencia estereotipada que empieza con un comportamiento de «protesta» y sigue con la «desesperación». La fase de protesta se caracteriza por sentimientos de pánico, vocalizaciones de angustia y comportamiento de búsqueda. El pánico suele combinarse con la rabia —«¡¿Dónde está?!»—, lo que genera otro conflicto, esta vez entre PÁNICO-DOLOR e IRA. Mientras una emoción hace que queramos tener cerca a la persona cuidadora, ahora y para siempre, al mismo tiempo la otra hace que queramos destruirla. La culpa, una emoción secundaria que inhibe la IRA, es el resultado aprendido típico. Es un buen ejemplo de cómo las emociones secundarias (como la culpa, la vergüenza, la envidia y los celos) se derivan de situaciones de conflicto. A diferencia de las emociones básicas, son constructos aprendidos, híbridos de emoción y cognición (como demuestra la investigación de Barrett).

La fase de desesperación se caracteriza por sentimientos de desesperanza, lo que se suele entender por «rendirse». La explicación convencional es que si los llantos y la búsqueda de la cría no conducen pronto al reencuentro, el precio que puede acabar pagando por alertar a sus depredadores respecto a su situación vulnerable empieza a ser mayor que los beneficios. Por otra parte, si la cría se aleja demasiado del hogar base, se reducen sus oportunidades de que la madre la encuentre a su regreso. Por tanto, en el balance estadístico, y como ocurría con la interrupción vasovagal, rendirse (por doloroso que sea) pasa a ser la estrategia de supervivencia heredada.

He aquí una descripción clásica de la cascada de separación en los niños humanos, realizada por el psicoanalista John Bowlby:

[La protesta…] puede empezar de inmediato o puede retrasarse; dura entre unas horas y una semana o más. Durante esta fase, el niño pequeño da señales de angustia profunda por haber perdido a su madre e intenta recuperarla mediante el pleno ejercicio de todos sus limitados recursos. En general llorará a gritos, sacudirá la cuna, dará vueltas y mirará impaciente hacia cualquier visión o sonido que pueda ser su madre desaparecida. Todo su comportamiento sugiere una firme expectativa de que la madre regresará. Mientras tanto, tenderá a rechazar todas las figuras alternativas que se ofrezcan a hacer cosas por él, si bien algunos niños se aferrarán desesperados a una niñera.

[La desesperación…] sigue a la protesta; la preocupación del niño por la desaparición de su madre sigue siendo evidente, aunque su comportamiento indica una creciente desesperanza. Los movimientos físicos activos se van reduciendo o se terminan y puede llorar de forma monótona o intermitente. Se muestra retraído e inactivo, no exige nada a las personas de su entorno y parece entrar en un estado de profundo duelo.[165]

Obviamente, este último estado recuerda a la depresión, que suele ir acompañada de culpa. Por ello Panksepp y otros (entre los que me incluyo) aplicaron su explicación de los mecanismos cerebrales de PÁNICO-DOLOR al desarrollo de nuevos tratamientos para los trastornos del estado de ánimo.[166] Químicamente, la transición de la «protesta» a la «desesperación» está mediada por unos péptidos llamados «opioides», que desactivan la dopamina (en el caso descrito más adelante, en la p. 146, se ven sus efectos). De ahí que la depresión

se caracterice por sentimientos opuestos a los que caracterizan a la BÚSQUEDA.[167] La trayectoria anatómica del componente PÁNICO de este sistema desciende desde la circunvolución del cíngulo anterior hasta la SGP, que es donde terminan todos los circuitos de la emoción.[168] (Más adelante explicaré por qué todos los ciclos afectivos acaban y también empiezan en la SGP del tronco encefálico).

Cabe señalar que este circuito mediado por los opioides evolucionó a partir del sistema analgésico más antiguo del cerebro; la angustia mental de pérdida es un desarrollo de los mecanismos corporales para el dolor sensorial.[169] Este es un buen ejemplo de la transición sin fisuras que existe en la naturaleza entre los afectos sensoriales que salvan la vida y los sentimientos emocionales. No hay nada «ficticio» en las emociones. Las sensaciones dolorosas asociadas a la separación y la pérdida —combinadas con el aprendizaje por experiencia— tienen una función causal para garantizar la supervivencia de los mamíferos y las aves, que necesitan cuidadores. Y además es algo aplicable pasada la infancia: los circuitos cerebrales que acabo de describir median en los vínculos de apego durante toda la vida, como por desgracia también median en muchas otras formas de adicción.

(6) CUIDADO es el otro lado del apego. Porque no solo necesitamos recibir cuidados y cariño, también necesitamos cuidar nosotros de los más pequeños, sobre todo de nuestra descendencia. El supuesto instinto maternal existe en todos, pero no en el mismo grado, porque está mediado por sustancias químicas con presencia más elevada (por término medio) en el sexo femenino: estrógeno, prolactina, progesterona y oxitocina, cuyos niveles se disparan durante el embarazo y el parto.[170] Aquí también cabe destacar cómo se solapan la química y los circuitos cerebrales de CUIDADO, PÁNICO-DOLOR y LUJURIA típica femenina.[171] Eso, por sí solo, ya podría explicar por qué la depresión es mucho más común (casi tres veces más) en las mujeres que en los hombres. Aproximadamente un 80 por ciento de las mujeres saben desde la infancia que es «bueno» acunar a los bebés hacia la izquierda de la línea media del cuerpo, mientras que los hombres tienden a descubrirlo (instintivamente) cuando son padres.[172] Por otro lado, hasta los chicos sin la menor experiencia saben en general qué hay que hacer cuando llora un bebé. No lo pinchan con el dedo ni lo agarran por el pie para ver si eso ayuda; saben (lo predicen de forma innata) que es «bueno» estrecharlos contra el cuerpo y mecerlos mientras se los calma con la voz.

Y sin embargo, como averigua toda madre o padre, con eso no basta.

Para criar a un bebé hasta que madura hace falta mucho más que instinto. En consecuencia, al igual que con las demás emociones, tenemos que aprender por experiencia qué hacer en las incontables situaciones imprevistas que surgen. En este sentido, y también como ocurre con las demás emociones, nuestras decisiones se guiarán por sentimientos (de cuidado y preocupación), que nos dicen si vamos bien o mal. Otra razón de que no baste con una pulsión de crianza es que no sentimos solamente amor hacia nuestros hijos, como confirmaría cualquier padre o madre. Los conflictos resultantes deben resolverse mediante procesos híbridos cognitivo-emocionales.

Aprender a conciliar de manera flexible las distintas necesidades emocionales entre sí constituye el fundamento de la salud mental y de la madurez. Pensemos por ejemplo en las relaciones románticas sostenibles, que requieren una integración sensata de la LUJURIA con un apego infantil de tipo PÁNICO-DOLOR (recordemos el complejo virgen-prostituta, que surge de la incapacidad de conciliar los sentimientos sexuales con los afectivos). También cuesta conciliar los vínculos afectivos con el sistema errante de BÚSQUEDA (pensemos en la emoción de la novedad), y con las inevitables frustraciones que provoca la IRA (de aquí que haya tantas peleas domésticas), que a su vez entra en conflicto con las preocupaciones del CUIDADO de la crianza, y así sucesivamente. Mantener relaciones duraderas es solo un ejemplo de los muchos desafíos a los que se enfrenta todo corazón humano. Las emociones nos sirven de brújula para gestionar los problemas de la vida. Son los sentimientos los que guían todo el aprendizaje a partir de la experiencia en las distintas formas que he desgranado. No obstante, la biología nos aporta otra pulsión para ayudarnos en el camino:

(7) JUEGO. Necesitamos jugar. Es el medio por el cual se reclaman y se defienden territorios, se forman jerarquías sociales y se forjan y mantienen los límites internos y externos de los grupos.

A menudo la gente se sorprende al saber que se trata de una pulsión biológica, pero, en la infancia, todos los mamíferos emprenden juegos bruscos y energéticos. Si se ven privados de su cuota diaria, intentarán compensarlo al día siguiente, como una recuperación. Todos sabemos qué clase de juegos físicos son, aunque la forma varía ligeramente de una especie de mamífero a otra. La sesión de juego empieza con alguna postura o gesto de «invitación»; si es aceptado, comienza el juego. El otro animal o niño sale en persecución del primero, hasta que se detienen y forcejean o se hacen cosquillas, turnándose para

estar encima o debajo y acompañándose de carcajadas o de la vocalización equivalente que corresponda a su especie (hasta las ratas «se ríen»).[173] Luego se reincorporan y se persiguen en dirección contraria. El estado emocional asociado también es universal: se llama «diversión».

A los niños les encanta jugar. Lo que no quita que en la práctica la mayoría de los episodios de juego acaben en lágrimas. Esto nos da una pista importante para saber de qué se trata, en términos biológicos: consiste en encontrar los límites de lo que es socialmente tolerable y permisible. Cuando el juego deja de ser divertido para uno de los que juegan, normalmente porque decide que el otro no está siendo «justo», dejan de jugar. Han alcanzado su límite. Marcar esos límites resulta crucial para la formación y conservación de grupos sociales estables. Y en especies sociales como la nuestra, la supervivencia de un grupo es importante para la supervivencia de cada uno de sus miembros.

Un criterio básico en este sentido es la dominancia. En cualquier situación de juego, uno de los participantes ocupa el papel principal y el otro el rol sumiso. Ambos se divierten, siempre y cuando el dominante no insista en serlo todas las veces. Al parecer, la proporción aceptable de turnos es más o menos de sesenta a cuarenta. La «regla de 60-40» de reciprocidad dice que el jugador sumiso sigue jugando mientras se le den suficientes oportunidades de ser también el dominante.

Esto revela una segunda función del JUEGO, a saber, el establecimiento de jerarquías sociales, un «orden de picoteo». Así, los juegos bruscos dejan paso (a través del desarrollo) a juegos más organizados y francamente competitivos. Y, por supuesto, el juego no se limita a la variedad de juegos bruscos. Los humanos tenemos los juegos de simulación, en los que los participantes prueban distintos roles sociales (por ejemplo, jugar a mamás y bebés, a maestros y alumnos, a médicos y pacientes, a policías y ladrones, a indios y vaqueros,[174] a reyes y siervos… Obsérvese la presencia constante de estatus y jerarquías de poder). No sabemos qué pasa por la imaginación de otros mamíferos cuando juegan, pero hay bastantes motivos para suponer que ellos también «prueban» diferentes roles sociales, y que con ello aprenden lo que pueden hacer y lo que no sin meterse en líos.

Esto sugiere una tercera función biológica para el JUEGO. Nos obliga a tener en cuenta —y nos condiciona para ello— los sentimientos de los demás. En caso contrario, se negarán a jugar con nosotros y perderemos el enorme placer que produce. Puede que el bruto de

turno quiera quedarse con todos los juguetes, pero perderá toda la diversión. Al parecer, por eso mismo evolucionó el JUEGO (y por eso se le atribuye tanto placer), porque fomenta las formaciones sociales viables. Es decir, es un gran vehículo para desarrollar empatía.[175]

Los episodios de juego se interrumpen de golpe cuando pierden su cualidad de simulación. Si encerramos a nuestra hermana pequeña y tiramos la llave, no solo hemos incumplido la regla de 60-40, sino que ya no estamos jugando al juego de policías y ladrones; solo estamos encerrando a nuestra hermana. Dicho de otro modo, lo que rige nuestro comportamiento mutuo ahora es el MIEDO o la IRA, en lugar del JUEGO. Lo mismo ocurre con los otros juegos enumerados. Jugar a médicos y pacientes es un juego hasta que se convierte en sexo real; entonces lo rige la LUJURIA. El hecho de que el JUEGO sobrevuele, por decirlo de alguna forma, todas las demás emociones instintivas —probándolas y aprendiendo sus límites— podría ser la razón de que no haya sido posible identificar un solo circuito cerebral para el JUEGO. Lo más probable es que los utilice todos.[176] En cualquier caso, quien dude de que jugar es en definitiva un instinto básico debería leer el maravilloso libro de Sergio y Vivien Pellis, The Playful Brain.[177]

No siempre nos gusta reconocer como un comportamiento natural de los seres humanos, similar al de otros mamíferos, la reclamación de territorios y el establecimiento de jerarquías sociales con normas claras. (Las normas que rigen el comportamiento de los primates son de una complejidad extraordinaria). La estructura de las familias, los clanes, los ejércitos e incluso las naciones —casi cualquier grupo social— es innegablemente jerárquica y territorial, y ha sido así a lo largo de la historia. Cuanto mayor es el estatus social de alguien dentro del grupo, más acceso tiene a los recursos del territorio que controla el grupo. Esta observación no es un asunto de preferencia personal; es una realidad. Si no hacemos frente a estas realidades, no podemos empezar a gestionarlas. El hecho de que existan pulsiones emocionales no significa que no tengamos control sobre ellas, que estemos obligados a inclinarnos ante «la ley de la selva», pero debemos ser conscientes del riesgo de ignorarlas.

Es fácil ver cómo el JUEGO, en concreto, genera normas sociales. Las normas regulan el comportamiento del grupo y, de esta forma, nos protegen de los excesos de nuestras necesidades individuales. También es fácil constatar cómo las normas sociales fomentan formas complejas de comunicación y, en consecuencia, cómo contribuyen a la emergencia del pensamiento simbólico. La cualidad de simulación del JUEGO sugiere que incluso podría ser el precursor biológico del

pensamiento en general (esto es, de la acción virtual frente a la acción real; véase más arriba). Algunos científicos también creen que soñar es el JUEGO nocturno, porque probamos emociones instintivas en un mundo de simulación. En este sentido cabe señalar que en el trastorno del comportamiento REM, en el que se ha perdido la parálisis motora que suele acompañar los sueños por culpa de una lesión del tronco encefálico, los pacientes (y los animales de experimentos) encarnan físicamente los distintos estereotipos instintivos, como la huida, la paralización, el embiste depredador y el ataque afectivo.

Espero que ahora se pueda apreciar que, mientras los afectos corporales tienen cierta intensidad e inmediatez que los hace parecer de alguna manera ineludibles, los afectos emocionales funcionan también a través del sentimiento consciente. Aunque no siempre los reconocemos por lo que son, regulan casi todo nuestro comportamiento voluntario a través de sus distintas estribaciones internas. El comportamiento voluntario consiste en esencia en tomar decisiones aquí y ahora. ¿Y cómo podemos tomar decisiones si no las basamos en algún sistema de evaluación que nos diga qué opción es mejor o peor? Estos valores son lo que los sentimientos aportan al comportamiento.

Sin embargo, como nuestros estados emocionales de fondo no siempre son reconocidos por el cerebro cognitivo, y por tanto no siempre se declaran de forma autorreflexiva, no vemos el patrón general hasta que echamos la vista atrás y atamos cabos. Cuando hablo de «nosotros», no me refiero únicamente a los legos, que tienen todo el derecho del mundo a desconocer los hechos experimentales. Por muy cierto que sea que enseguida tendemos a ignorar el papel fundamental que los sentimientos tienen en la vida cotidiana, a quien en realidad me refiero es a la corriente principal de la ciencia cognitiva contemporánea. Los científicos cognitivos se han acostumbrado a ignorar los sentimientos.

Y, en cambio, como demostrará el siguiente capítulo desde una perspectiva neurológica, cualquier relato científico de la conciencia que ignore este papel fundamental de los sentimientos se estará perdiendo lo más importante.

[146] Alboni, 2014.

[147] El siguiente párrafo parafrasea a Merker, 2007.

[148] Véase Nummenmaa et al., 2018 —un artículo muy interesante—, que respalda varias de las conclusiones a las que llegaré a continuación, como la relativa a la imposibilidad de afectos «neutros», la materialización de los afectos y la naturaleza categórica de los afectos.

[149] Por ejemplo, el miedo conlleva respiración acelerada, aumento de la frecuencia cardíaca, redireccionamiento de la sangre desde los intestinos hacia la musculatura esquelética y, por todo ello, un estado tónico de alerta y de preparación para escapar. Si la sensación de miedo conllevara respiración lenta, disminución de la frecuencia cardíaca, redireccionamiento de la sangre desde los músculos esqueléticos hacia los intestinos y, por todo ello, una actitud destonificada, no sería miedo. La naturaleza materializada de los afectos queda evidenciada en los escáneres térmicos de cuerpo entero de las distintas emociones (véase Nummenmaa et al., 2018). Véase también Niedenthal, 2007.

[150] Estas cuestiones filosóficas se tratan con más detalle en el capítulo 11.

[151] Los urólogos llaman a esta situación «urgencia de la llave en la cerradura».

[152] Ekman et al., 1987.

[153] Barrett, 2017.

[154] Véanse los detalles empíricos y las referencias bibliográficas en Panksepp, 1998, y Panksepp y Biven, 2012.

[155] Véase LeVay, 1993.

[156] Para los interesados en los detalles anatómicos, señalo que en el sexo masculino típico, el centro del circuito de LUJURIA es el hipotálamo anterior (especialmente los núcleos intersticiales), desde donde desciende a través del núcleo del lecho de la estría terminal hasta la SGP. Desde el punto de vista químico, la hormona esteroidea testosterona (liberada por los testículos y que actúa en gran medida sobre el hipotálamo anterior) media la liberación en el cerebro de un péptido denominado vasopresina, responsable de la excitación sexual masculina. En el sexo femenino típico, el hipotálamo ventromedial es

el lugar de control sexual, y las principales sustancias químicas son el estrógeno y la progesterona (ambas liberadas por los ovarios, el equivalente femenino de los testículos). Estas hormonas median la actividad de la oxitocina en el cerebro, un péptido que gobierna gran parte de la respuesta típica del sexo femenino. La LUJURIA también está mediada por otros péptidos, como la hormona liberadora de gonadotropina y la colecistoquinina.

[157] La BÚSQUEDA también está mediada por el neurotransmisor glutamato y una serie de péptidos, como la oxitocina, la neurotensina y la orexina.

[158] Comentaré con más detalle el tema del pensamiento en el capítulo 10.

[159] Tiene su origen en las partes mediales de la amígdala y pasa por el núcleo del lecho de la estría terminal y las secciones medial y perifornical del hipotálamo en su camino hacia la SGP. Su neuromodulador dominante es un péptido llamado sustancia P, que actúa junto con el glutamato y la acetilcolina. Este último hecho podría explicar por qué el trastorno del comportamiento REM se expresa tantas veces a través de la IRA.

[160] El circuito del MIEDO surge de la amígdala central y basolateral. Químicamente está mediado por el neurotransmisor glutamato, además de los péptidos: inhibidor de la unión a diazepam, corticoliberina, colecistoquinina, alfa melanotropina y neuropéptido Y.

[161] Véase Tranel et al., 2006, que se centró expresamente en la emocionalidad subjetiva de S. M.

[162] Véase Blake et al., 2019.

[163] Lo mismo se aplica a otras tareas similares, como la «intuición» en la tarea de juego Iowa (Turnbull et al., 2014). Como expliqué a Nicholas Humphrey, que objetó mi uso de la expresión en inglés «gut feeling» (corazonada o presentimiento) en lugar de «guessing» (suposición o conjetura) para hablar de la visión ciega, todo depende de las preguntas que nos hagamos (conferencia «The Science of Consciousness», Interlaken, junio de 2019).

[164] LeDoux, 1996.

[165] Bowlby, 1969.

[166] Por ejemplo, Yovell et al., 2016; y Coenen et al., 2019.

[167] Solms y Panksepp, 2010.

[168] A través del núcleo del lecho de la estría terminal, el área preóptica y el tálamo dorsomedial. Además de los mecanismos opioides descritos en el texto, el PÁNICO está mediado por el neurotransmisor glutamato y los neuropéptidos oxitocina, prolactina y corticoliberina.

[169] Posiblemente, esta es la razón por la que el dolor mental se somatiza tantas veces como dolor físico. Véanse Eisenberger, 2012; y Tossani, 2013.

[170] El CUIDADO también está mediado por la dopamina.

[171] El circuito CUIDADO desciende desde el cíngulo anterior a través del núcleo del lecho de la estría terminal, el área preóptica y el área tegmental ventral, hasta la SGP.

[172] Forrester et al., 2018.

[173] Panksepp y Burgdorf, 2003. Véase también: www.youtube.com/watch?v=j-admRGFVNM.

[174] Por fortuna, ya no es un juego popular. Mi mención no supone en modo alguno un apoyo al hecho por otra parte irrefutable de la opresión de los nativos norteamericanos por los colonos.

[175] La empatía se deriva de la postura intencional, o «teoría de la mente», que por supuesto no está desarrollada a igual nivel en todas las especies de mamíferos. Por lo tanto, el desarrollo de la empatía no es en ningún momento un proceso automático, como podría sugerir la teoría de la «neurona espejo». La empatía no es un reflejo; es un logro del desarrollo (véase Solms, 2017a).

[176] Sin embargo, el tálamo dorsomedial y el área parafascicular parecen ser especialmente importantes, al igual que la SGP, por supuesto. Si hay un modulador dominante del JUEGO son los opioides mu, pero esto podría reflejar simplemente el hecho de que la seguridad (es decir, bajo PÁNICO-DOLOR) es una condición previa necesaria para el JUEGO. Otros posibles moduladores del JUEGO son el glutamato, la acetilcolina y los cannabinoides. El circuito talamocortical identificado por Zhou *et al.* (2017) solamente corresponde a un aspecto específico del JUEGO, a saber, la dominancia. Véanse también Van der Westhuizen y Solms, 2015; Van

der Westhuizen et al., 2017.

[177] Pellis y Pellis, 2009.

06

La fuente

Un supuesto básico de la neuropsicología —basado en el método clínico-anatómico— es que si una función mental concreta la realiza una región cerebral concreta, la lesión total de esa región resultará en la pérdida total de esa función. Como he demostrado, en lo que respecta a la conciencia, la corteza cerebral no cumple con ese supuesto. Y aún menos si acudimos a la teoría cortical de la conciencia, porque hay lesiones en otros lugares del cerebro que la suprimen por completo, incluso tratándose de lesiones muy pequeñas.

Hace más de setenta años, los fisiólogos Giuseppe Moruzzi y Horace Magoun demostraron que los gatos pierden la conciencia con solo practicar minúsculas incisiones que desconectan la corteza de la formación «reticular» (en forma de red) del tronco encefálico.[178] Esta formación debe de tener aproximadamente unos 525 millones de años, porque la poseen todos los vertebrados, desde los peces hasta los humanos. Desde que Moruzzi y Magoun lo descubrieron, los investigadores han confirmado en especies de todo tipo que lesiones relativamente pequeñas en esta formación —conocida técnicamente como el «sistema reticular activador»— provocan el coma. No hace mucho, por ejemplo, David Fischer y sus colegas identificaron en pacientes humanos con apoplejía del tronco encefálico una diminuta región de dos milímetros cúbicos «propia del coma» en el tegmento mesopontino superior (véase fig. 1).[179]

Hay dos explicaciones posibles. La primera es que esta formación tan densa y enredada del tronco del encéfalo es el punto del que surge la conciencia: es el manantial oculto de la mente, la fuente de su esencia. Yo mantengo esta teoría, como hizo Jaak Panksepp. La segunda explicación es que aquí ocurre como con el cable de un televisor: es necesario, pero no suficiente y apenas ayuda a entender cómo funciona el aparato. Esta es la teoría imperante.

Supongamos que es cierta la segunda opción. Entonces cabría esperar que la conciencia pudiera encenderse o apagarse mediante la estimulación del tronco encefálico. Como mucho, podemos atenuarla, al igual que una bajada de tensión apaga progresivamente la pantalla del televisor. No esperaríamos que el aparato pudiera restablecer sobre la marcha lo que se está emitiendo en ese momento. Y, sin

embargo, un electrodo implantado en una formación reticular del tronco encefálico de una mujer de sesenta y cinco años (para el tratamiento de la enfermedad de Parkinson) provocó fiablemente esta sorprendente respuesta:

En tan solo cinco segundos, el rostro de la paciente expresó una tristeza profunda [...]. Aunque seguía alerta, la paciente se inclinó a la derecha, empezó a llorar y comunicó verbalmente sentimientos de tristeza, culpa, inutilidad y desesperanza, como «siento que me derrumbo, ya no quiero vivir, ni ver nada, ni oír nada, ni sentir nada...». Cuando se le preguntó por qué lloraba y si sentía dolor, respondió: «No, estoy harta de la vida, ya he tenido bastante... No quiero vivir más, estoy asqueada de la vida... Todo es inútil, no valgo para nada, este mundo me da miedo». Cuando se le preguntó por qué estaba triste, contestó: «Estoy cansada. Quiero esconderme en un rincón... Me quejo de mí misma, claro... No tengo remedio, no sé por qué le molesto». [...] La depresión desapareció menos de 90 segundos después de interrumpir la estimulación. Durante los cinco minutos siguientes, la paciente presentó un estado ligeramente hipomaníaco, riendo y gastándole bromas al examinador, tirándole de la corbata. Luego recordó todo el episodio. La estimulación [en otro lugar del cerebro, que era la diana real del electrodo] no provocó esta respuesta psiquiátrica.[180]

Esta paciente no tenía antecedentes de síntomas psiquiátricos de ningún tipo.

Lo mismo ocurre con la estimulación química o el bloqueo de estos núcleos centrales del tronco encefálico. La mayoría de los antidepresivos —potenciadores de la serotonina— actúan sobre neuronas cuyos cuerpos celulares están situados en una región del sistema reticular activador llamada «núcleos del rafe» (véase fig. 1).

Por decirlo de alguna manera, son la «fuente» de la serotonina. Los antipsicóticos —bloqueantes dopaminérgicos— actúan sobre neuronas originadas en otra parte del sistema reticular activador: el área tegmental ventral (véase fig. 2). Lo mismo ocurre con los ansiolíticos, muchos de los cuales bloquean una sustancia química llamada noradrenalina, producida por neuronas originadas en el complejo del locus cerúleo (también fig. 1), otra parte del sistema reticular activador. Todas estas neuronas están agrupadas en la formación

reticular del tronco encefálico. Los psiquiatras no harían probaturas con esta región del cerebro si se limitara a encender y apagar la conciencia, en cuyo caso solo les interesaría a los anestesistas. En consecuencia, la segunda teoría tiene que ser errónea.

Las neuroimágenes funcionales del cerebro en estados emocionales apuntan a la misma conclusión. La tomografía por emisión de positrones durante los estados de DOLOR, BÚSQUEDA, IRA y MIEDO, por ejemplo, muestra que la mayor actividad metabólica se produce en el núcleo del tronco encefálico (y otras regiones subcorticales; véase fig. 9, p. 149), mientras que la corteza presenta desactivación. Las resonancias magnéticas funcionales durante el orgasmo revelan lo mismo: la actividad hemodinámica que se correlaciona con este estado intensamente afectivo se localiza casi de forma exclusiva en el mesencéfalo.[181]

Tanto los estudios sobre lesiones como la estimulación cerebral profunda, la manipulación farmacológica y las neuroimágenes funcionales apuntan a la misma conclusión: la formación reticular del tronco encefálico genera afecto. Así pues, aparentemente la única parte del cerebro que sabemos que es necesaria para despertar la conciencia en su conjunto tiene una influencia igual de fuerte sobre otra función mental: sentir. En el capítulo anterior ya he explicado cómo los sentimientos impregnan toda la experiencia consciente: con independencia de las otras funciones que pueda albergar, una de las tareas centrales de la conciencia es tener y gestionar sentimientos (que se originan dentro de nosotros y regulan nuestras necesidades biológicas). Y ahora resulta que las fuentes neurológicas del afecto y de la conciencia están, como mínimo, muy entrelazadas, eso si no son la misma maquinaria. Contrariamente a la visión empirista clásica, según la cual la conciencia fluye a través de nuestros sentidos, y contrariamente a la afirmación que cité de Meynert, que estaba basada en esa visión, parece que el cerebro sí que «irradia su propio calor».

¿Cómo deberíamos llamar al medio básico, a esta misteriosa materia mental que parece brotar de nuestro interior? No podemos llamarlo «estado de vigilia», como hizo Zeman, porque eso nos obligaría a describir el soñar como un tipo de vigilia, lo cual es absurdo. Tampoco podemos llamarlo «nivel» cuantitativo, como hicieron Moruzzi y Magoun, porque los hechos que acabamos de describir demuestran que implica rasgos intensamente cualitativos.

Intentemos, pues, utilizar el tercer término empleado en las publicaciones: excitación (arousal). A mí me parece una palabra neutra y conveniente. Tanto la vigilia como el soñar implican

excitación; y no excluye la calidad, como sí que hace la palabra nivel. De hecho, excitación sugiere positivamente sentimiento.

Pero ¿qué es la excitación? Hasta ahora la hemos comentado en términos directamente conductuales, como en el caso de las distinciones entre coma, estado vegetativo y vigilia plenamente receptiva. Por lo general se mide con la escala de coma de Glasgow, que mide las respuestas de apertura ocular de los pacientes, sus respuestas verbales a preguntas y sus respuestas motoras a instrucciones y al dolor. Sin embargo, también puede definirse fisiológicamente.

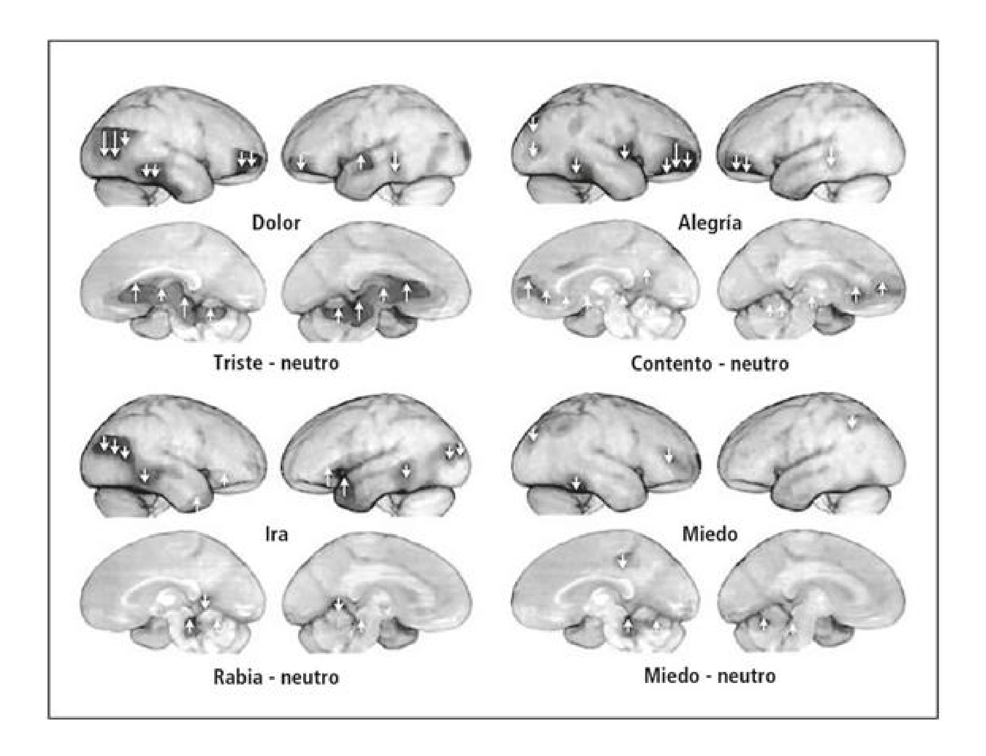


Figura 9

. Tomografías por emisión de positrones de cuatro estados emocionales (cortesía de Antonio Damasio). Las flechas ascendentes indican regiones de activación incrementada, y las flechas descendentes, regiones de activación disminuida. La zona resaltada en la imagen de «Alegría» parece mostrar la activación del sistema de BÚSQUEDA.

Un electroencefalograma genera trazados gráficos de la actividad eléctrica cortical. Por sí sola —es decir, si se desconecta del sistema reticular activador, incluso cuando procesa aportes sensoriales—, la corteza produce el patrón de ondas delta, una serie de ondas de gran amplitud que aparecen unas dos veces por segundo (es decir, una frecuencia de 2 Hz). Cuando es estimulada por el sistema reticular activador sin que haya aportes sensoriales, la corteza suele producir el ritmo theta (4-7 Hz)[182] o el ritmo alfa (ondas desincronizadas a frecuencias de 8-13 Hz, donde «desincronizadas» significa erráticas). Cuando está activa procesando información externa, la corteza suele mostrar el patrón beta (ondas desincronizadas de muy baja amplitud con frecuencias de 14-24 Hz) o gamma (ondas de baja amplitud con frecuencias muy altas de 25-100 Hz). Gamma es el ritmo que se suele asociar más con la conciencia.

En la actualidad también se puede medir la excitación fisiológica mediante neuroimágenes funcionales, que reproducen literalmente la actividad cerebral cartografiando patrones regionales de cambio en las tasas metabólicas. La figura 3 ilustraba esta técnica, con referencia a las distintas fases del sueño. La fila inferior recoge la excitación del sueño REM, que suele asociarse —aunque, como ya he demostrado, no en exclusiva— a la conciencia durante el sueño y que se origina en la parte superior del tronco encefálico. Las neuroimágenes de algunos de los estados emocionales básicos de la conciencia antes descritos (véase fig. 9) y del orgasmo muestran lo mismo: la excitación del tronco del encéfalo.

La corteza solo adquiere conciencia en la medida en que es excitada por el tronco encefálico. La relación entre ambos es jerárquica: la conciencia cortical depende de la excitación del tronco. Por ello el patrón de ondas delta mostrado en la fila superior de la figura 10 —que no está asociado con el comportamiento consciente— es generado por la activación cortical intrínseca, y también por ello el ritmo gamma mostrado en la fila inferior de la misma figura —que está fuertemente asociado a la conciencia— puede ser impulsado por el sistema reticular activador por sí solo.[183] Esto también explica por qué la disminución de la excitación fisiológica mostrada en la fila superior de la figura 3 coincide con la mengua de la conciencia y la aparición del sueño, en tanto que el aumento de excitación mostrado en la fila inferior coincide con la reaparición de la conciencia durante el sueño. Estos datos no son discutibles.

Veamos un poco más de cerca los verdaderos mecanismos cerebrales implicados. Aunque antes conviene establecer una distinción básica

entre las dos formas en las que las neuronas se comunican entre sí. Una distinción que adquiere especial importancia para la conciencia.

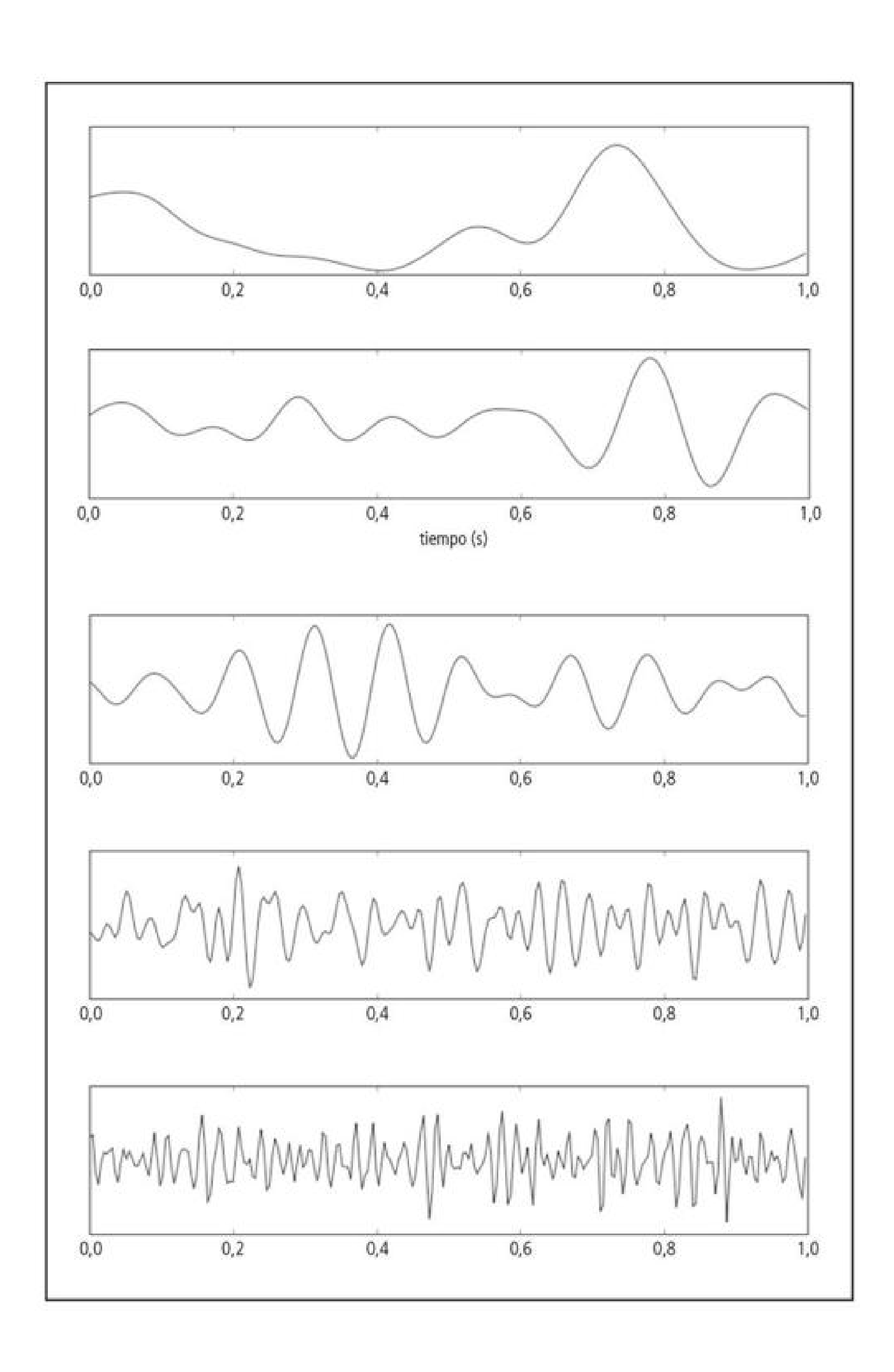


Figura 10

. Patrones típicos de actividad cortical electroencefalográfica. El patrón menos excitado (delta) se muestra en la parte superior, y el patrón más excitado (gamma) en la parte inferior. Los patrones intermedios, de arriba abajo, son theta, alfa y beta.

Quien sienta un mínimo interés por el cerebro sabrá que las neuronas transmiten mensajes a través de redes muy complejas. A este proceso se lo llama «transmisión sináptica», porque se envían mensajes a través de las sinapsis, que son las estructuras mediante las cuales una neurona manda señales a otra (la palabra sinapsis viene del griego y significa «unión»). La transmisión sináptica se realiza por medio de unas moléculas llamadas «neurotransmisores», que pasan de una neurona a la siguiente, excitando o inhibiendo a la neurona postsináptica, según sea la molécula en cuestión (el glutamato y el aspartato son neurotransmisores excitadores, y el ácido gamma-aminobutírico [GABA] es un neurotransmisor inhibidor). Si la siguiente neurona es excitada por una ráfaga de neurotransmisores, pasa sus moléculas a las neuronas que le siguen en la red. En caso contrario, no. A partir de ahí, las moléculas transmisoras se degradan enseguida o son reabsorbidas en la neurona presináptica para limitar la duración de su efecto, en un proceso llamado «recaptación».

La transmisión sináptica está enfocada a un objetivo, es binaria (sí/no) y es rápida. Es el aspecto del funcionamiento cerebral que más recuerda a la computación digital, lo que explicaría por qué les ha interesado siempre tanto a los neurocientíficos de orientación computacional. Tiene lugar en todo el sistema nervioso, incluida la corteza cerebral; pero no es intrínsecamente consciente. Es decir, este tipo de neurotransmisión se produce en la corteza tanto si es consciente como si no. Y no tiene prácticamente nada que ver con la excitación.

Lo que no es tan sabido es que la transmisión sináptica tiene lugar bajo la influencia constante de un proceso fisiológico completamente distinto. A ese otro tipo de actividad neuronal se le llama «modulación postsináptica» y, a diferencia de la transmisión sináptica, es caótica, ineludiblemente química y no tiene nada que ver con lo que ocurre en un ordenador. Surge de forma endógena del sistema reticular activador (y de otras estructuras subcorticales e incluso de algunas estructuras corporales no neurológicas); y tiene todo que ver con la excitación.

Las piezas centrales de este proceso son un tipo de moléculas llamadas «neuromoduladores». A diferencia de los neurotransmisores, estas moléculas se esparcen de forma difusa por el encéfalo, esto es, son liberadas más o menos cerca de agrupaciones enteras de neuronas y no tanto en sinapsis individuales. (Cabe señalar que algunas moléculas neuromoduladoras también pueden actuar como neurotransmisores). En vez de pasar mensajes por «canales» concretos, se extienden por secciones enteras de la red y así regulan el «estado» global de la corteza. Por ejemplo, la corteza está en un estado distinto en las filas superior e inferior de la figura 3 (sueño de ondas lentas vs. sueño REM) y en los cuatro estados emocionales mostrados en la figura 9 (DOLOR vs. BÚSQUEDA vs. IRA vs. MIEDO). En cada uno de estos estados, procesa la información de distinta forma. Así, si alguien nos llama mientras dormimos, reaccionamos de manera muy distinta a si nos llaman cuando estamos completamente despiertos.[184] Algo parecido ocurre cuando se nos acerca una persona desconocida: nuestra reacción cambiará mucho si estamos en un estado de BÚSQUEDA que si estamos en un estado de MIEDO. En el primer caso tal vez saludemos a la persona o incluso entablemos una conversación, mientras que en el segundo puede que apartemos la mirada con la esperanza de que no nos vea.

Esta distinción entre «canal» y «estado» es una buena pista para entender las dos formas en las que las neuronas se comunican entre sí.[185] El estado de la corteza afecta a la intensidad variable con la que pasa el mensaje por sus canales; por decirlo de alguna manera, el estado ajusta lo «alto que se hablan» los distintos canales entre sí (véase fig. 11, p. 155). Esto explica por qué el mismo sonido —en nuestro ejemplo, alguien llamándonos— se esparce por toda la corteza durante la vigilia pero se queda aislado en la corteza auditiva durante el sueño, y por qué una persona desconocida excita una red cerebral durante el estado de MIEDO y otra durante el estado de BÚSQUEDA.

Este es el quid de lo que llamamos «excitación». No obstante, hay que tener en cuenta que la excitación cortical se puede modular al alza y a la baja en la transmisión, hasta el punto de suprimirla por completo, como sucede cada noche cuando nos vamos a dormir (por eso algunos fisiólogos prefieren hablar de «modulación» más que de «excitación»). Así pues, la excitación determina qué impulsos sinápticos se transmitirán, y con qué intensidad, como en el ejemplo de la reacción a que nos llamen durante el sueño frente a la vigilia.

La transmisión sináptica es binaria (encendido/apagado, sí/no, 1/0), pero la neuromodulación postsináptica gradúa la probabilidad de que un conjunto dado de neuronas dispare. Hace variar las posibilidades

estadísticas de que ocurra algo en ellas. Este ajuste analógico y probabilístico de las tasas de disparo lo efectúan unos receptores ubicados en distintos puntos a lo largo de la neurona. A diferencia de los neurotransmisores, los neuromoduladores tienen efectos relativamente lentos y duraderos, no solo porque las propias sustancias químicas persisten, sino porque es bastante frecuente que los canales que disparan necesiten tiempo para predisponerse al disparo. Si se estimula alguna parte de la red, quedará estimulada hasta el momento de su modulación a la baja. Esto influye en la plasticidad neuronal y explica mucho cómo funciona el aprendizaje. Los estados de excitación hacen que lo que aprendemos quede más inscrito en los canales de nuestro cerebro. Así, por ejemplo, es más probable que recordemos un viaje en el que buscamos ansiosos un destino desconocido que cuando viajamos al mismo lugar por costumbre, con el piloto automático.

¿De dónde proceden los neuromoduladores? Proceden de todo el cuerpo, incluidas las glándulas pituitaria, suprarrenales, tiroidea y sexuales (que producen diversas hormonas) y el hipotálamo (que produce innumerables péptidos). Pero si hablamos del cerebro, la fuente central de «excitación» es el sistema reticular activador. La excitación reticular del tronco encefálico libera los cinco neuromoduladores más conocidos: la dopamina (que se origina principalmente en el área tegmental ventral y la sustancia negra), la noradrenalina (que se origina principalmente en el complejo del locus cerúleo), la acetilcolina (que se origina principalmente en el tegmento mesopontino y los núcleos basales del prosencéfalo), la serotonina (que se origina principalmente en los núcleos del rafe) y la histamina (que se origina principalmente en el hipotálamo tuberomamilar). Ya hemos visto de cerca algunas de estas sustancias químicas en capítulos anteriores, y no por casualidad. Cada una de las moléculas y sus receptores asociados, de los que existen muchos subtipos, es responsable de un aspecto diferente de la excitación moduladora. Además de estas cinco sustancias, existen muchas otras, principalmente hormonas y péptidos de acción lenta (de los que hay más de cien en el cerebro), que modulan sistemas neuronales muy específicos.

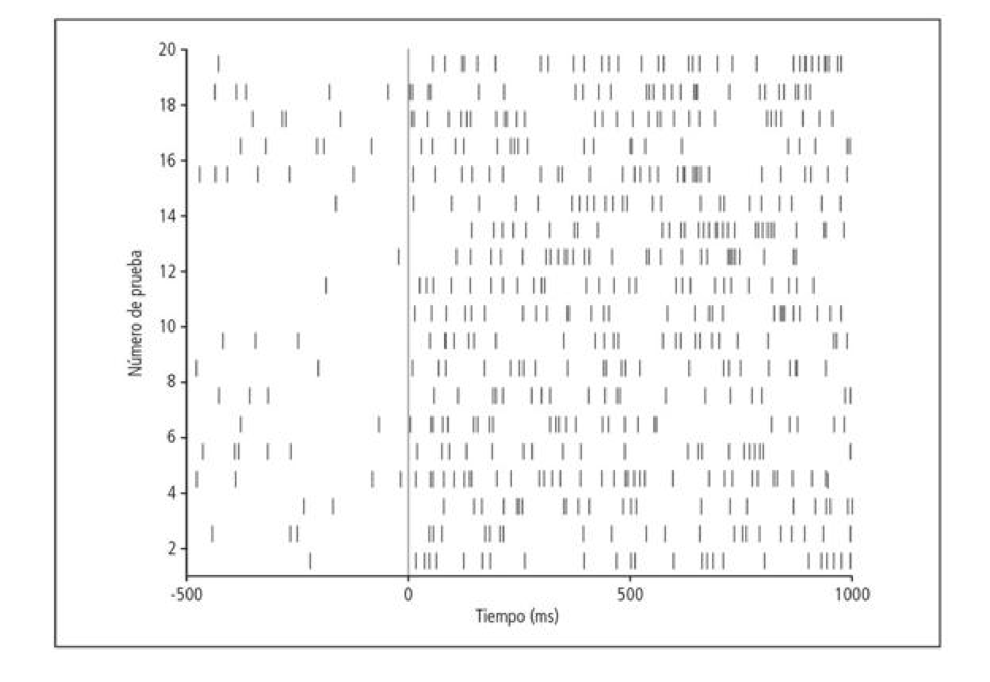


Figura 11

. La imagen es un gráfico de los trenes de impulsos nerviosos de 20 neuronas (representados en el eje de ordenadas) durante un periodo de 1,5 segundos (representado en el eje de abscisas) durante la presentación de un estímulo visual. La presentación del estímulo se produce en el tiempo 0 del eje de abscisas (representado por la segunda línea vertical). En este momento, las neuronas, que tienen una tasa de disparo de referencia de 6 Hz (por término medio, cuando no hay estímulo), aumentan su tasa de disparo media a 30 Hz. Un tren de impulsos nerviosos es una secuencia en la que una neurona dispara (= impulsos nerviosos) y no dispara (= silencios). Puede representarse como una secuencia digital de información: «1» para un impulso nervioso y «0» para un silencio. Por ejemplo, un tren de impulsos nerviosos codificado podría leerse como «001111101101». Los dos primeros ceros representan aquí el tiempo de latencia entre la presentación del estímulo y el primer impulso. Es importante señalar que las tasas de disparo no están determinadas únicamente por el estímulo, sino que surgen de una interacción entre el estímulo, la potenciación o inhibición a largo plazo de las neuronas (es decir, si el estímulo es familiar o desconocido) y su nivel actual de modulación. El ajuste de la modulación postsináptica es el efecto de la excitación sobre las neuronas. Por esta razón, el estímulo representado aquí podría no provocar respuesta alguna de las mismas neuronas cuando

la excitación se modula a la baja.

El efecto de todos estos moduladores se ve determinado por la presencia o ausencia de los receptores y subtipos de receptores pertinentes. En otras palabras, aunque los neuromoduladores estén esparcidos por todo el encéfalo, solo influyen en las células que contienen los receptores pertinentes. No podemos pensar en la excitación como en un proceso sin pulir. Al contrario, es muy polifacética y se despliega en muchas dimensiones, tanto espaciales como temporales. Los mensajes potenciales que llevan los neuromoduladores flotan por las numerosas redes cerebrales, pero, como los analgésicos, solo se utilizan cuando hace falta. Por ejemplo, mientras leen esto, sus sistemas sensoriales registran estímulos de fondo de toda índole sin que les presten atención, pero si son madres y su bebé recién nacido empieza a llorar, esto anulará su concentración en el libro y les hará ser conscientes de inmediato del bebé. La causa es el incremento de los niveles de ciertas hormonas y péptidos —estrógeno, progesterona, prolactina y oxitocina— que flotan por su cerebro cuando han tenido un bebé y que cambian su estado (activan el sistema cerebral de CUIDADO).

Los neuromoduladores solo pueden modular (al alza o a la baja) señales que realmente existen, es decir, canales activos en ese momento. Su liberación es difusa, pero solo influyen en las neuronas que (1) contienen los receptores pertinentes y (2) están justo entonces activas. El mismo modulador tiene efectos distintos en clases de receptores distintos de lugares distintos. Y la especificidad aumenta aún más en el caso de los moduladores que no son liberados únicamente en el tronco del encéfalo (y el sistema endocrino), sino también en los propios circuitos neuronales. Los cinco moduladores de la excitación más conocidos que he citado antes se originan sobre todo en el tronco encefálico reticular. Las hormonas se originan en otra parte del cuerpo y llegan al cerebro principalmente a través del torrente sanguíneo. Los más específicos son los péptidos, muchos de los cuales se originan en el hipotálamo. Cada péptido hace algo distinto, también aquí en función del tipo de receptor y de su ubicación. Muchos de ellos, que modulan las emociones básicas de maneras muy especializadas, quedan recogidos en las notas del capítulo 5.

En lo que llevamos de capítulo he abordado dos preguntas: ¿de dónde

procede anatómicamente la excitación? y ¿cómo se despierta fisiológicamente? Las respuestas son que se genera sobre todo, aunque no en exclusiva, en el tronco encefálico y en el hipotálamo, y que excita el prosencéfalo modulando la neurotransmisión. Si volvemos al capítulo 5, veremos que la razón por la que se genera la excitación es que responde a las demandas de trabajo endógenas que se hacen a la mente. Estas demandas adoptan la forma de todo tipo de señales de «error» que convergen en el centro del cerebro. Gran parte de ellas se gestionan de manera automática e inconsciente, y solo exigen respuestas conscientes en situaciones en las que no basta con las respuestas automáticas. Esto me lleva a la pregunta principal de este capítulo: ¿dónde se produce este salto aparentemente mágico del reflejo automático al sentimiento volitivo?

Lo que ahora describiré y explicaré no es definitivo, porque los propios datos todavía no están del todo claros. Sin embargo, el conocimiento neurocientífico actual está lo bastante avanzado como para permitirnos trazar un esbozo de la situación general. El pionero de esta investigación fue Jaak Panksepp, seguido por Antonio Damasio. Damasio tiene el don de saber ver el panorama completo, pero el neurocientífico al que creo que más debemos en este sentido, en lo relativo a cómo el cerebro da el salto misterioso desde el reflejo autónomo a la acción voluntaria, es Björn Merker. Merker ha estudiado la excitación del tronco del encéfalo y los mecanismos de orientación en un amplio abanico de especies de vertebrados, incluidas las aves, los roedores, los felinos y los primates.

Al parecer, el salto de la vigilia vegetativa a la excitación afectiva depende de la integridad de un reducido y denso nudo de neuronas que rodean el canal central del mesencéfalo, la sustancia gris periacueductal (SGP), donde convergen todos los circuitos afectivos del encéfalo (véase fig. 6 para la ubicación de esta estructura). Por eso los pacientes con lesiones localizadas en la SGP «se quedan mirando al vacío psicoafectivo»:[186]

Una lesión importante de la SGP [causa] un deterioro espectacular de todas las actividades conscientes. […] Por ejemplo, los primeros estudios en los que se introdujeron electrodos lesionantes desde el cuarto ventrículo por el acueducto y hasta el borde caudal del diencéfalo mostraron déficits sorprendentes en la conciencia de gatos y monos, operacionalizados por su incapacidad para presentar cualquier comportamiento intencional aparente y su falta global de respuesta a estímulos emocionales.[187] Aunque las lesiones en otras

áreas superiores del cerebro pueden dañar las «herramientas [cognitivas] de la conciencia», normalmente no afectan a la base de la intencionalidad en sí misma. Sin embargo, las lesiones de la SGP, sí, y aunque la destrucción absoluta de tejido cerebral sea mínima.[188]

El estado que acabo de describir es el «estado vegetativo». Se diferencia del coma en que conserva la vigilia. Sin embargo, el ciclo circadiano de sueño y vigilia no es más que otra función autónoma.[189] Por eso al estado vegetativo también se lo llama «vigilia sin respuesta», un aparente oxímoron que revela la importante distinción entre vigilia (vegetativa) y excitación (afectiva) (lo que Panksepp llama «intencionalidad» en la cita anterior).[190] También por esto prefiero el término excitación a vigilia o nivel de conciencia. La palabra excitación acoge —e incluso sugiere positivamente— la respuesta emocional y la intencionalidad, que, como volvemos a ver aquí, residen en el núcleo del comportamiento consciente.[191] Esto es lo que la SGP añade al funcionamiento vegetativo y automático.

¿Y cómo lo hace? La SGP no forma parte del sistema reticular activador, aunque esté justo a su lado y densamente interconectada con él.[192] La principal diferencia entre estos núcleos y la SGP es la dirección del flujo de información entre ellos y el prosencéfalo. Mientras el sistema reticular activador ejerce sobre todo su influencia hacia arriba en la corteza, la corteza solo transmite señales devolviéndolas hacia abajo a la SGP.

La SGP es el punto de reunión final de todos los circuitos de afecto del encéfalo. Por tanto, mientras el prosencéfalo es excitado por el sistema reticular activador, la SGP es excitada (por así decirlo) por el prosencéfalo. Podríamos ver el sistema reticular activador y la SGP, respectivamente, como el origen y el destino de la excitación del prosencéfalo.

En consecuencia, la SGP se conceptualiza como el final de la red «descendente» para el afecto, en contraste con las redes de afecto «ascendentes» y «moduladoras» de los núcleos de control corporal del tronco encefálico y del sistema reticular activador.[193] Esto significa que la SGP es el principal centro de salida de todos los circuitos afectivos, canalizando la información hacia los efectores musculoesqueléticos y viscerales que generan la «emoción propiamente dicha».[194] (Aquí cito directamente estudios serios, para asegurarme de que describo con precisión nuestra comprensión actual de estas funciones). La red descendente para el afecto está

«implicada en acciones motoras específicas invocadas en las emociones, así como en el control de la frecuencia cardíaca, la respiración, la vocalización y la conducta de apareamiento».[195] El papel de la SGP en esta red es actuar como «interfaz de los estímulos salientes entre el prosencéfalo y el tronco encefálico inferior».[196] En este sentido, la SGP se conceptualiza como un centro para «equilibrar o transicionar información relacionada con la saliencia de la supervivencia»,[197] esto es, la SGP funciona «orquestando distintas estrategias de afrontamiento cuando se expone a estresores externos».[198] «Ofrece un punto de reunión masivo de los sistemas neuronales que generan emocionalidad».[199] Por lo tanto, desempeña el papel central en «la defensa homeostática de la respuesta del individuo, integrando la información aferente de la periferia y la información de los centros superiores».[200]

Para decirlo lisa y llanamente: todos los circuitos afectivos convergen en la SGP, que es el principal centro de salida para los sentimientos y los comportamientos emocionales. Por este motivo, «intensidades inferiores de estimulación eléctrica de esta zona cerebral excitarán a los animales a más tipos de acciones emocionales coordinadas que la estimulación de cualquier otro punto del encéfalo».[201]

A simple vista, las columnas poco diferenciadas que constituyen la SGP rodean el canal central del tronco encefálico a lo largo de catorce milímetros. El canal central, por el que circula un líquido incoloro (el líquido cefalorraquídeo), es el acueducto cerebral o «acueducto de Silvio», antes llamado así por el anatomista del siglo XVII que lo describió por primera vez. Su ubicación en el centro del mesencéfalo da nombre a la SGP: «sustancia gris periacueductal» significa simplemente «materia gris que rodea el canal».[202] Este núcleo primitivo del cerebro es realmente lo que Freud describió como «su interior más recóndito». Se divide en dos grupos de columnas funcionales.[203] Una de ellas, la posterior, es para las «estrategias de afrontamiento» activas o los comportamientos defensivos, como las reacciones de lucha-huida, el aumento de la tensión arterial y el alivio del dolor sin opioides.[204] Aquí es donde terminan los circuitos de MIEDO, IRA y PÁNICO-DOLOR. La columna delantera corresponde a las estrategias pasivas de afrontamiento-defensa, como quedarnos paralizados con hiporreactividad, el «comportamiento de enfermedad» duradero, la bajada de la tensión arterial y el alivio del dolor con opioides.[205] Los circuitos de LUJURIA, CUIDADO y BÚSQUEDA terminan en esta columna delantera.

La SGP es el camino común final hacia el producto afectivo, lo que hace que le corresponda a ella, en una palabra, «elegir» qué toca hacer

a continuación, una vez que los diversos circuitos afectivos y sus comportamientos condicionados asociados hayan hecho sus aportaciones a la acción.[206] Y para decidir tiene que evaluar las señales de error residuales que le llegan vía los sistemas de afecto, sopesando sus distintas propuestas en función de los imperativos biológicos definitivos de sobrevivir y reproducirse, mientras cada señal de error le comunica su necesidad constitutiva. En resumen, la SGP tiene que establecer prioridades para la siguiente secuencia de acciones.

Sin embargo, las prioridades de acción no se pueden determinar únicamente en función de las necesidades. Hay que contextualizar las necesidades, no solo entre ellas sino también respecto a las oportunidades existentes en ese momento.

Volvamos al ejemplo del control respiratorio. Pensemos primero en el contexto interno: si siento alarma por asfixia y sed al mismo tiempo, la alarma por asfixia tiene prioridad por delante de la sed, pero, comparada con la tristeza, por ejemplo, la sed es la que tendría prioridad. Consideremos ahora el contexto externo: si experimento alarma por asfixia, el contexto me dirá si necesito —por ejemplo— eliminar una obstrucción de las vías respiratorias o salir de una habitación llena de dióxido de carbono. El contexto externo en el que actúo conserva su relevancia a medida que se suceden los acontecimientos y yo decido una y otra vez qué toca hacer a continuación.

Si recordamos mi definición de afecto en el capítulo 5, veremos que incluye muchas cosas que no suelen verse como afectivas. Así, por ejemplo, la alarma por asfixia y la sed no son sensaciones meramente corporales; también transmiten —y no menos que otros sentimientos, como la ansiedad por separación— valores intrínsecos (bondad y maldad biológicas). El dolor es un buen ejemplo para distinguir entre la valencia intrínseca del afecto y su contexto exteroceptivo. La sensación desagradable del dolor es lo que lo hace afectivo, en tanto que la sensación somática exteroceptiva transmite la localización de un estímulo doloroso: «dolor procedente de la mano izquierda». No es una distinción filosófica; los aspectos duales se pueden manipular por separado, por ejemplo, mediante la estimulación de la SGP y de la corteza parietal, respectivamente. Esto hace posible que en algunas situaciones clínicas pueda uno percibir que su mano izquierda ha sido pinchada con un alfiler sin sentir el menor dolor.[207]

Con este trasfondo, Björn Merker hace una profunda observación.[208] Aunque la SGP se halla anatómicamente debajo de la corteza, su

importancia en términos funcionales es primordial. Después de que la corteza y otras estructuras del prosencéfalo hayan realizado su trabajo cognitivo —después de que hayan hecho sus aportaciones a la acción —, la decisión final sobre «qué toca hacer a continuación» se efectúa a nivel del mesencéfalo. Dichas decisiones adoptan la forma de sensaciones afectivas, generadas por la SGP, que pueden anteponerse a cualquier estrategia cognitiva formulada durante la secuencia de acciones anterior. La SGP elige en cada momento el afecto que determinará y modulará la siguiente secuencia. Por ejemplo, en mi primer día de trabajo como neuropsicólogo, decidí cognitivamente volver a las camas de los pacientes para leer sus historiales, pero lo que ocurrió en realidad afectivamente fue que me desmayé.

Ya sabemos, claro, que la vida no es una serie interminable de emergencias. La mayor parte del tiempo, lo que prevalece es la actividad de BÚSQUEDA por defecto (interés y compromiso de fondo). Pero, volviendo a las emergencias, no podemos sentir a la vez las múltiples señales de error que llegan al mesencéfalo transmitiendo necesidad, porque no podemos hacerlo todo a la vez. Y aquí está el meollo de la función de priorización de la SGP: ¿cuál de esas señales de error es la más saliente, dadas las circunstancias del momento? ¿Cuál de mis problemas se puede aplazar (o gestionar automáticamente) y cuál siento que debo resolver?

La exigencia de tener en cuenta las circunstancias del momento requiere, naturalmente, la intervención de más elementos. La SGP emite su veredicto con la ayuda de una estructura adyacente del mesencéfalo, los tubérculos cuadrigéminos superiores (véase fig. 6). Se hallan justo detrás de la SGP y están divididos en varias capas, cada una de las cuales proporciona cartografías simples derivadas de los sentidos corporales. Las capas más profundas suministran mapas motores del cuerpo, mientras que las capas superficiales son responsables de los mapas espaciosensoriales. Entre unas y otras reúnen una representación increíblemente comprimida e integrada del mundo exteroceptivo, procedente en parte de la corteza, pero también de zonas sensoriomotoras subcorticales como el nervio óptico (véase también fig. 6). Así, los tubérculos cuadrigéminos superiores representan de forma sintetizada el estado del cuerpo objetivo (sensorial y motor) de cada momento, más o menos como la SGP controla su estado subjetivo (necesidad). Merker llama «triángulo decisorio» del cerebro a esta interfaz afectiva-sensorial-motora entre la SGP, los tubérculos cuadrigéminos superiores y la región locomotora del mesencéfalo.[209] Panksepp lo denominó el «YO primordial», la fuente misma de nuestro ser sintiente.[210]

Así pues, las decisiones del mesencéfalo sobre qué toca hacer a continuación se basan en la retroalimentación de los circuitos afectivos del cerebro junto con sus mapas sensoriomotores, cada uno de los cuales nos va actualizando sobre distintos aspectos de «en qué punto estamos». Recordemos mi ejemplo de cuando me di cuenta de pronto de mi necesidad de orinar al término de una conferencia de dos horas. En aquel momento, al percibir la oportunidad, mi necesidad corporal fue sentida y se transformó en una pulsión volitiva. En resumen, como el triángulo decisorio del mesencéfalo toma en cuenta las condiciones tanto internas como externas, prioriza las opciones de comportamiento basándose no solo en las necesidades del momento, sino también en las oportunidades del momento.

La capa más profunda de los tubérculos cuadrigéminos superiores consiste en un mapa que controla los movimientos oculares. Este mapa es intrínsecamente más estable que los mapas sensoriales superpuestos, porque estos se calibran tomando el primero de referencia, estableciendo así el «punto de vista» unificado que caracteriza la experiencia perceptiva subjetiva. Esto es lo que permite que nos experimentemos en escenas visuales estables con independencia de lo deprisa que se muevan nuestros ojos, unas tres veces por segundo. La escena estabilizada también nos dice que lo que percibimos solo es eso, una escena, una perspectiva construida sobre la realidad, no la realidad en sí. Por esta razón nos experimentamos como si viviéramos dentro de nuestra cabeza.[211] En el capítulo 10 explicaré con más detalle la naturaleza virtual de la percepción.

Como hemos visto cuando hablábamos de la visión ciega, el «mapa tipo pantalla bidimensional» —como lo llama Merker— del mundo sensoriomotor que generan los tubérculos cuadrigéminos superiores es inconsciente en los seres humanos.[212] Contiene poco más que una representación de la dirección de la «desviación del objetivo» —donde el objetivo es el centro de cada ciclo de acción— para producir la orientación de la mirada, la atención y la acción. Brian White lo llama mapa de «saliencia» o «prioridad». Panksepp explica que así es como nuestras «desviaciones respecto de un estado de reposo acaban siendo representadas como estados de disposición para la acción».[213] Yo no sabría decirlo mejor.

La conciencia perceptiva del mundo que nos rodea es posible gracias a la ayuda de una corteza convenientemente excitada, de la que carecen —a diferencia de la conciencia afectiva— los niños hidranencefálicos y los animales decorticados. Los tubérculos cuadrigéminos superiores proporcionan cartografías condensadas inmediatas de objetivos y acciones potenciales, pero la corteza proporciona «representaciones»

detalladas que usamos para guiar cada secuencia de acciones a medida que se desarrolla. Además de estas imágenes tan diferenciadas, en el prosencéfalo subcortical hay muchos programas de acción inconsciente, llamados «procedimientos» y «respuestas», no imágenes. (Pensemos, por ejemplo, en los tipos automatizados de recuerdos en los que confiamos para ir en bicicleta o para seguir una ruta hasta un destino conocido). Están codificados sobre todo en los núcleos basales subcorticales, la amígdala y el cerebelo. Los recuerdos no son meros registros del pasado. En términos biológicos se refieren al pasado, pero están ahí para el futuro. Todos ellos son, en su esencia, predicciones destinadas a cubrir nuestras necesidades. Trataré este importante aspecto en los siguientes capítulos.

Las tendencias motoras que se activan a través de la selección de afectos del prosencéfalo despiertan reflejos e instintos simples; y eso es todo lo que hacen en los bebés, los niños hidranencefálicos y muchos animales. Sin embargo, como ya sabrán, esos comportamientos automáticos se van sometiendo a control individualizado durante el desarrollo mediante el aprendizaje por experiencia. De esta forma, las respuestas estereotipadas se complementan con un repertorio más flexible de opciones. La secuencia conductual derivada de cada nuevo ciclo de acciones se despliega en sentido ascendente sobre estos niveles cada vez más amplios de control prosencefálico, desde las «respuestas» procedimentales hasta las «imágenes de la memoria» figurativas. Esto genera lo que Merker denomina «un mundo tridimensional panorámico y completamente articulado compuesto por objetos sólidos y con forma: el mundo de nuestra experiencia fenoménica conocida».[214]

Recordemos que los recuerdos a largo plazo sirven para el futuro. Una vez que el triángulo decisorio del mesencéfalo ha evaluado la retroalimentación comprimida que le llega de cada acción anterior, lo que activa es un proceso ampliado de prealimentación que se despliega en dirección opuesta, a través de los sistemas de memoria del prosencéfalo, hasta generar un contexto previsible para la secuencia motora elegida. Este es el producto de todo nuestro aprendizaje. Cuando una necesidad nos empuja al mundo, no descubrimos por primera vez ese mundo con cada nuevo ciclo. Activa nuestro conjunto de predicciones sobre las probables consecuencias sensoriales de nuestras acciones, basándose en nuestra experiencia previa de cómo cubrir la necesidad elegida en las circunstancias del momento.

A partir de ahí, la acción voluntaria conlleva un proceso de contrastar nuestras expectativas con las consecuencias reales de nuestras

acciones. La comparación produce una señal de error, con la que reevaluamos nuestras expectativas sobre la marcha y ajustamos nuestros planes de acción en consecuencia. En eso consiste precisamente el comportamiento «voluntario»: en decidir qué hacer en condiciones de incertidumbre, mediados por las consecuencias de cada acción. Y hablamos de consecuencias sentidas afectiva y perceptivamente, por lo que las señales de error residuales tanto afectivas como sensoriomotoras convergen en el triángulo decisorio del mesencéfalo. A mí me gusta la expresión que aplica Jakob Hohwy al proceso mental que controla el comportamiento voluntario: «predecir el presente».[215]

Para tener una impresión clínica de lo que estoy explicando, recuerden lo que le ocurrió al señor S., el ingeniero eléctrico con psicosis de Korsakoff cuyo caso describí en el capítulo 2. Cuando miró por mi ventana, esperaba ver Johannesburgo. Y esto se debía a su trastorno de la memoria: no había actualizado su modelo predictivo conforme a los sucesos recientes. Cuando el paisaje nevado de Londres con el que se encontró no cubrió sus expectativas, no cambió de opinión. Ignoró la señal de error que entraba y se aferró a su predicción original, diciendo: «No, no. Yo sé que estoy en Johannesburgo. Que uno esté comiendo pizza no significa que esté en Italia». En otras palabras, su cerebro moduló a la baja la señal de error. Por tanto, no ajustó su programa de acción, algo destinado a acabar mal, porque no pudo aprender de la experiencia. Si no hubiera tenido a la familia y los médicos atendiendo todas sus necesidades, con toda probabilidad habría muerto.[216]

Poco se reconoce que las percepciones de aquí y ahora están guiadas en todo momento por predicciones, generadas sobre todo desde la memoria a largo plazo. Pero es así. Y es lo que explica que sean muchas menos las neuronas que propagan las señales desde los órganos de los sentidos externos hasta los sistemas de memoria internos que en sentido contrario.[217] Por ejemplo, la proporción entre conexiones que entran y conexiones que salen en el cuerpo geniculado lateral —que transmite información desde los ojos hacia la corteza visual y viceversa; véase fig. 6— es de aproximadamente una a diez. El trabajo pesado lo realizan las señales predictivas que se encuentran con las sensoriales que llegan desde la periferia. Esto ahorra una cantidad ingente de procesamiento de la información y, con él, de trabajo metabólico. Teniendo en cuenta que el encéfalo consume un 20 por ciento de nuestros suministros de energía totales, la eficiencia conseguida es importante. ¿Por qué tratar todo lo que hay en el mundo como si nunca lo hubiéramos visto antes? El cerebro prefiere propagar hacia el interior solo la parte de la información

entrante que no se ajusta a sus expectativas. Por este motivo hay quien ahora describe la percepción como «fantasía» y «alucinación controlada»; empieza con un guion previsible que luego se ajusta para responder a la señal entrante.[218] En este sentido, los anatomistas clásicos tenían razón: el procesamiento cortical consiste sobre todo en la activación de las «imágenes de la memoria», convenientemente reordenadas para predecir el siguiente ciclo de percepción y acción.

Lo que estoy diciendo es aplicable a todos los vertebrados, no solo a los humanos. Puede que hasta sea aplicable a todos los organismos dotados de cerebro, o de sistema nervioso. (Los insectos, por ejemplo, tienen estructuras cerebrales que funcionan en gran medida como nuestro sistema reticular activador).[219] Ahora sabemos que la forma fundamental de la conciencia es el afecto, que nos permite resolver situaciones imprevistas «sintiendo». Pero ¿cómo se transforma el afecto endógeno en exterocepción consciente?

Panksepp sugiere que los «afectos sensoriales» (p. ej., dolor, asco y sorpresa) podrían haber forjado un puente evolutivo entre esos dos aspectos de la experiencia. Los afectos sensoriales son a la vez sensaciones internas y percepciones externas; son percepciones inherentemente valenciadas matizadas por sentimientos concretos. Así, por ejemplo, el dolor se siente de forma distinta al asco, y respondemos a ellos también de forma distinta, retirándonos o con arcadas, según cuál sea. Según Panksepp, la conciencia se ha extendido, pasando por este puente evolutivo, hasta la percepción en general, porque contextualiza el afecto. Después de todo, el mundo exterior solo adquiere valor para nosotros porque tenemos que cubrir nuestras necesidades en él; adquiere valor en relación con el afecto.

Así pues, mi respuesta a la pregunta de por qué sentimos la percepción, la acción y la cognición es que las sentimos porque contextualizan el afecto. Es como si nuestra experiencia perceptiva dijera: «Me siento así sobre eso».[220] La percepción es, por así decirlo, incertidumbre aplicada. Es razonable entonces decir, como hace Merker, que las conciencias afectiva y perceptiva utilizan una «divisa común»; son distintas clases de sentimiento, pero no dejan de ser sentimiento.[221] Aquí solo hablo de percepción consciente. Y basta entonces con dar un pequeño paso para inferir que nuestras cinco modalidades de conciencia perceptiva (vista, oído, tacto, gusto y olfato)[222] evolucionaron para matizar las distintas categorías de información externa que registran nuestros órganos sensoriales,[223] de la misma manera en que las siete clases de conciencia afectiva

comentadas en el capítulo anterior matizan las distintas categorías de necesidad emocional en los mamíferos.

Sin embargo, es importante recordar también que la percepción cortical es apercepción. Lo que aparece en la conciencia no son las señales sensoriales en bruto tal y como llegan transmitidas desde la periferia, sino inferencias predictivas derivadas de huellas mnémicas de dichas señales y sus consecuencias. El hecho de que las sensaciones se esparzan, por así decirlo, sobre las inferencias corticales estabilizadas para crear lo que yo llamo «sólidos mentales»[224] —el mundo exterior tal y como se manifiesta en la percepción— explica en parte las distintas cualidades fenoménicas del afecto frente a la percepción consciente. Volveré a tratar estos temas en el capítulo 10.

El modelo interno del cerebro es el mapa que utilizamos para navegar por el mundo y, claramente, para generar un mundo previsible. Pero no podemos dar por buenas todas nuestras predicciones. De hecho, hay dos aspectos que considerar en el «contexto previsible» que genera el modelo interno: por un lado tenemos el contenido real de nuestras predicciones, y por otro, nuestro nivel de confianza en su exactitud. Como todas las predicciones son probabilísticas, también habrá que codificar el grado de incertidumbre prevista asociado a ellas. Las propias predicciones son suministradas por las redes de memoria a largo plazo del prosencéfalo, que filtran el presente a través de la lente del pasado. Pero la segunda dimensión, el ajuste de los niveles de confianza, es la función central de la excitación moduladora.

Esto es lo que se torció en el caso del señor S.: depositó demasiada confianza en sus predicciones. O, dicho de otro modo, a consecuencia de su estado no sopesó bien sus señales de error (no les prestó suficiente atención).[225] ¿Cómo pudo ser? En cierta forma, esto ya lo hemos visto, pero desde el otro lado. Modular las señales neuronales significa únicamente ajustar su intensidad aumentándola o disminuyéndola. Eso es lo que hace el triángulo decisorio del mesencéfalo cuando elige qué señales va a potenciar: modula la intensidad de la señal a través del sistema reticular activador. La intensidad de la señal de referencia representa el contexto previsible. Sin embargo, conforme se despliega el contexto real (experimentado), hay que ajustar la intensidad de las señales. En consecuencia, confiar más en una señal de error implica necesariamente confiar menos en la predicción que ha llevado a ese error. Si se nos dispara el detector de humos pero no parece que se esté quemando nada, falla uno de los dos: o el detector o las apariencias; pero la discrepancia exige un

veredicto. Esto es lo que ocurre en el triángulo decisorio del mesencéfalo: se valoran las propuestas competidoras y se le concede la victoria a una. El resultado de este proceso se transmite al prosencéfalo vía el sistema reticular activador, que actúa basándose en nuestras expectativas y después —conforme tiene lugar la acción elegida— libera nubes de moléculas neuromoduladoras para ajustar las señales en nuestras redes de memoria a largo plazo, regulando al alza los canales del prosencéfalo en donde se almacenan algunas predicciones y regulando otros a la baja. Esto lleva a su vez al aprendizaje a través de la experiencia. De esta forma, seguimos mejorando nuestro modelo generativo del mundo gracias al método de ensayo y error.

Todo lo que hacemos en el ámbito de la incertidumbre se guía por estos niveles de confianza fluctuantes. Creemos que sabemos lo que ocurrirá si actuamos de determinada manera, pero ¿lo sabemos realmente? Si nuestra convicción desciende por debajo de cierto umbral, no actuaremos o cambiaremos de estrategia. En la esfera exteroceptiva todo va bien mientras las cosas salen según lo previsto, y mal cuando se impone la incertidumbre. En consecuencia, nos sentimos bien o mal: el aumento de confianza (en una predicción) es bueno y su reducción es mala. Y eso nos lleva a intentar reducir al mínimo la incertidumbre en nuestras expectativas.

Como hemos visto antes, la percepción y la acción son un proceso progresivo de prueba de hipótesis, en el que el cerebro intenta una y otra vez suprimir las señales de error y confirmar sus hipótesis. (Precisamente la ciencia experimental es tan solo una versión sistematizada de este proceso cotidiano, como hemos visto en el capítulo 1: «Si la hipótesis X es correcta, debería producirse Y cuando yo haga Z»). Cuanto más se confirman nuestras hipótesis, más confiados estamos y menos excitación —menos conciencia— necesitamos. Podemos automatizar nuestras secuencias de acciones y dejarnos llevar por el modo por defecto. Pero si nos encontramos ante una situación imprevista —en la que nuestro modelo predictivo no nos arroja ninguna luz fiable—, las consecuencias de nuestras acciones adquieren mucha saliencia. Apagamos el piloto automático y nos volvemos hiperconscientes: el triángulo decisorio ajusta con minuciosidad nuestras predicciones mientras nos abrimos paso sintiendo las consecuencias de nuestras acciones y tomamos nuevas decisiones.

De esta forma, el sentimiento sigue siendo el factor común de toda conciencia, tanto afectiva como cognitiva. Su función es evaluar el éxito o fracaso de nuestros programas de acción y sus contextos

asociados. Pero no todos los fracasos tienen la misma importancia. Los sentimientos más intensos tienen que ver con incertidumbres de valor para la supervivencia, más que con incertidumbres vinculadas a eventualidades relativamente menores. Una mala predicción del tipo «ahora el coche al que sigo va a girar a la izquierda» puede tratarse en niveles periféricos de la jerarquía predictiva, en la que se tolera más el error, y provocar así cambios de lo que llamamos «atención» en lugar del afecto (si bien la atención también está modulada por el tronco encefálico reticular). Sin embargo, si de pronto parece que el coche al que seguíamos nos va a implicar en un accidente, es más probable que nuestro error excite una respuesta afectiva. La atención se puede dirigir y se puede captar. En este segundo caso irá acompañada de sentimientos de sobresalto o temor, que reescribirán sobre la marcha las predicciones fallidas para garantizar que nuestro modelo del mundo no vuelva a provocarnos la misma sorpresa desagradable.

Así es como funciona la conciencia en el cerebro en términos fisiológicos. Puede que no les resulte muy satisfactorio, porque, para empezar, apenas he considerado los abordajes filosóficos del problema mente-cuerpo. No he tratado la relación entre lo que algunos filósofos llaman sus «aspectos duales»: ¿por qué la fisiología objetiva de la conciencia se acompaña de una sensación fenoménica subjetiva? Cuando los científicos observan una correlación regular como esta entre dos conjuntos de datos cualesquiera, buscan una única causa subyacente. Quieren explicar por qué esos fenómenos co-ocurren. Por lo tanto, para entender las conjunciones regulares entre los aspectos fisiológicos y los aspectos psicológicos de la conciencia, tenemos que profundizar. Debemos superar las limitaciones de la psicología y la fisiología como disciplinas. Y eso los científicos solemos hacerlo recurriendo no a la metafísica, sino a la física.[226] En este caso, las respuestas que buscamos las hallaremos en el concepto físico de la entropía.[227]

[178] Moruzzi y Magoun, 1949.

[179] Fischer et al., 2016. La lesión se hallaba cerca (justo encima) del núcleo parabraquial medial. Curiosamente, es la región en la que Hobson identificó las células colinérgicas de origen del sueño REM. Véanse también Parvizi y Damasio, 2003; y Golaszewski, 2016.

[180] Blomstedt et al., 2008. El electrodo se colocó en la sustancia negra. El lugar previsto era el núcleo subtalámico.

[181] Damasio et al., 2000; Holstege et al., 2003.

[182] Aquí hablamos de la convexidad neocortical, no del hipocampo.

[183] Garcia-Rill, 2017.

[184] Holeckova et al., 2006.

[185] Este uso de los términos canal y estado es atribuible a Mesulam, 2000. En el capítulo 9 utilizo la analogía de los «modos de funcionamiento» para las funciones de «estado». Para una amena explicación de cómo se ve realmente el cerebro en diferentes estados o modos de funcionamiento a nivel celular, véase Abbott, 2020.

[186] Panksepp, 1998, p. 314. La cita de Panksepp continúa en el párrafo que sigue e incorpora las dos notas siguientes.

[187] Bailey y Davis, 1942.

[188] Depaulis y Bendler, 1991.

[189] Para una explicación muy amena al respecto, véase Walker, 2017. Hasta los moluscos y los equinodermos (como las estrellas de mar) muestran un ciclo de sueño-vigilia. Panksepp (1998, p. 135) señala que la regulación del sueño es filogenéticamente más antigua que el sistema reticular activador. Y a partir de ahí, formula una sugerencia intrigante: «En un principio, lo que ahora es el mecanismo del sueño REM mediaba la excitación selectiva de la emocionalidad. Es posible que, antes de la aparición de estrategias cognitivas complejas, los animales generasen la mayor parte de su comportamiento a partir de rutinas psicoconductuales de procesos primarios que ahora reconocemos como sistemas emocionales primitivos […] En otras palabras, muchos de los comportamientos de los animales antiguos pueden haber surgido en gran medida de subrutinas emocionales preprogramadas. Con el tiempo, aquellas soluciones conductuales sencillas fueron sustituidas por planteamientos cognitivos más sofisticados que requerían no solo más neocorteza, sino también nuevos mecanismos de excitación para mantener funciones de vigilia eficientes dentro de aquellas áreas cerebrales emergentes».

A la luz de lo que propongo más adelante, la sugerencia de Panksepp podría reformularse así; el prosencéfalo añade a los programas

motores instintivos inferiores y automatizados la capacidad de modular contextualmente el comportamiento emocional y, de esta forma, de aprender por experiencia.

[190] No debe equipararse al concepto filosófico de intencionalidad o «dotación de contenido» (aboutness). Panksepp se refiere a algo similar a la «volición», aunque más adelante veremos, cuando aborde el concepto filosófico, que está profundamente relacionado con la volición.

[191] Así, cuando Pfaff (2005) cubre exhaustivamente el tema de la excitación, que describe como «la fuerza más fundamental del sistema nervioso», operacionaliza el término como sigue: «La "excitación generalizada" es mayor en un animal o ser humano que está: (S) más alerta a estímulos sensoriales de todo tipo, (M) más activo motrizmente y (E) más reactivo emocionalmente». Dada la importancia que tiene el tema de la «excitación» en este libro, en el apéndice de la p. 367 cito un extenso extracto de Pfaff, 2005, que sirve también para llevarnos al tema del siguiente capítulo. Agradezco esta oportunidad de reconocer el trabajo seminal de Donald Pfaff, quien (ya a principios de los años noventa, cuando lo conocí) mostró una inusual apreciación de la formulación freudiana de «pulsión».

[192] La SGP proyecta hacia todos los núcleos fuente neuromoduladores del sistema reticular activador. Los otros destinos principales de las proyecciones de la SGP en el tronco encefálico son el hipotálamo medial, el núcleo cuneiforme, la formación reticular pontina, el núcleo solitario, el núcleo grácil, el núcleo reticular dorsal y la médula ventrolateral. Véase Linnman *et al.*

[193] Venkatraman, Edlow e Immordino-Yang, 2017. No me gusta la palabra descendente en este contexto, porque la SGP integra la retroalimentación afectiva cerebral superior y visceral inferior. Solo «desciende» en el sentido de que produce una salida motora. Convendría más decir «centrípeta», para poder contrastarlo con una red «centrífuga» (es decir, con lo que Edlow llama la red «moduladora»). Una red «centrípeta» incluiría tanto la red «descendente» como la «ascendente». Linnman *et al.* (2012) describen la SGP como el lugar de interacción entre los sistemas «límbico descendente» y «sensorial ascendente».

Nota: a lo largo de este libro, utilizo las palabras arriba y abajo, superior e inferior, ascendente y descendente, etc., no como juicios de valor, sino como localizadores anatómicos. A diferencia de otros órganos corporales, el cerebro es estructuralmente jerárquico. Está

estratificado como un yacimiento arqueológico, con los niveles más antiguos cubiertos por los más recientes. De ahí el título de este libro: El manantial oculto. El núcleo más profundo del tronco encefálico contiene las estructuras más antiguas, en términos evolutivos, y los niveles más altos de la corteza contienen las más recientes. Esto no significa que las estructuras inferiores (y más antiguas) sean menos importantes que las superiores (más recientes). Al contrario, desde el punto de vista funcional, las estructuras más altas del prosencéfalo no son más que elaboraciones de las más bajas del tronco encefálico.

[194] Venkatraman, Edlow e Immordino-Yang, 2017.

[195] Ibid.

[196] Linnman et al., 2012, p. 517; la cursiva es mía.

[197] Ibid.; la cursiva es mía. En estudios de imagen del cerebro humano se ha visto, y no es sorprendente, que la SGP pertenece a una «red de saliencia» (Seeley et al., 2007).

[198] Ezra et al., 2015, p. 3468; la cursiva es mía.

[199] Panksepp y Biven, 2012, p. 413; la cursiva es mía.

[200] Linnman et al., 2012, p. 506; la cursiva es mía.

[201] Panksepp y Biven, 2012.

[202] Antes se le llamaba «gris central».

[203] Las de «atrás» son la SGP lateral y dorsolateral. La de «delante» es la SGP ventrolateral. Esta clasificación no tiene en cuenta la SGP dorsomedial.

[204] Véase Venkatraman, Edlow e Immordino-Yang, 2017: «Cuando se estimula, esta columna produce vocalización emocional, confrontación, agresión y activación simpática, mostrada por el incremento de la tensión arterial, la frecuencia cardíaca y la respiración. […] Dentro de esta columna dorsolateral/lateral propiamente dicha hay dos partes. La parte rostral es responsable del poder/dominio (produciendo una respuesta de lucha), mientras que la parte caudal invoca el miedo (produciendo una respuesta de huida) con flujo sanguíneo hacia las extremidades».

[205] La columna delantera «recibe señales de dolor somático y visceral “lento y ardiente” mal localizadas y, al ser estimulada,

produce un afrontamiento pasivo, un comportamiento de enfermedad a largo plazo, paralización con hiporreactividad y una inhibición del flujo de salida simpático […] De este modo, es probable que participe en emociones de fondo como las que contribuyen al estado de ánimo» (ibid.).

[206] Simone Motta y colaboradores lo expresan así (Motta, Carobrez y Canteras, 2017, p. 39): «[La SGP] ha sido reconocida comúnmente como un lugar descendente en las redes neuronales para la expresión de diversos comportamientos, y se cree que proporciona respuestas estereotipadas. Sin embargo, cada vez hay más datos que sugieren que la SGP puede ejercer una modulación más compleja de varias respuestas conductuales y trabajar como un centro único de suministro de tono emocional primario para influir en los puntos prosencefálicos que median respuestas aversivas y apetitivas complejas».

[207] La SGP es un punto habitual de estimulación cerebral profunda para el tratamiento del dolor crónico, pero no disminuye las capacidades somatosensoriales corticales.

[208] Merker, 2007. Él, a su vez, atribuye esta idea a Penfield y Jasper, 1954.

[209] A efectos de este libro, utilizo una versión simplificada de la expresión técnica «triángulo de selección mesodiencefálico». Además, utilizo la expresión de Merker refiriéndome más a una interfaz de decisión que a un triángulo (es decir, una conexión entre necesidad y contexto). Como veremos en el capítulo siguiente, la acción («selección de acciones» para Merker) y la percepción («selección de objetivos» para Merker) —que juntas conforman el contexto— son dos caras de la misma moneda. Merker lo describe así (2007, p. 70): «Por mucho que el telencéfalo se haya expandido posteriormente, incluso hasta el punto de enterrar el mesodiencéfalo bajo una neocorteza de vertiginoso crecimiento en los mamíferos, nunca fue necesaria otra disposición, y por una razón muy simple. Ningún nervio eferente tiene su núcleo motor situado por encima del nivel del mesencéfalo. Esto significa que por el punto del tronco encefálico donde se unen el mesencéfalo y el diencéfalo, de sección transversal muy estrecha, […] pasa absolutamente toda la información necesaria para que el prosencéfalo pueda generar, controlar o influir en cualquier tipo de comportamiento». Merker llama a esta sección transversal «cuello de botella sinencefálico». Y añade: «Basta entonces con saber que en el cerebro de los vertebrados las motoneuronas más rostrales están situadas por debajo del cuello de botella sinencefálico para entender

que todo el contenido informativo del prosencéfalo debe someterse a una enorme reducción de datos en el curso de su traducción en tiempo real en comportamiento».

El hecho de que la decisión de qué toca hacer a continuación se tome a este nivel (tronco del encéfalo) —esto es, después de que las regiones del prosencéfalo hayan presentado sus distintas «ofertas»— queda perfectamente ilustrado por el ejemplo de la alarma por asfixia antes comentado: todas las consideraciones cognitivas quedan anuladas por la sensación de disnea, que se desencadena a nivel del tronco encefálico. Los centros de control respiratorio se hallan en la protuberancia y el bulbo raquídeo.

[210] Una década antes que Merker, Panksepp (1998, p. 312) describió como sigue la disposición funcional del YO: «Las capas más profundas de los tubérculos cuadrigéminos constituyen una cartografía motora básica del cuerpo [objetivo], que interactúa no solo con los sistemas visual, auditivo, vestibular y somatosensorial, sino también con circuitos emocionales cercanos de la SGP. La SGP elabora un mapa diferente, de tipo visceral, del cuerpo [subjetivo] junto con representaciones neuronales básicas de los sistemas de dolor, miedo, rabia, ansiedad por separación, comportamiento sexual y comportamiento maternal [como se resume en el capítulo anterior de este libro]. Junto a la SGP se halla la región locomotora mesencefálica, capaz de instigar patrones neuronales que deberían ser un sustrato esencial para la configuración de distintas tendencias de acción coherentes». Damasio y Carvalho (2013) también propusieron esta disposición funcional general para lo que denominaron el «protoyó». Señalaron que el complejo parabraquial y el núcleo del fascículo solitario proporcionan otros mapas sensoriales heteromodales primitivos del cuerpo. Estos mapas podrían ser precursores evolutivos de la función integradora que desempeña el triángulo decisorio del mesencéfalo.

[211] Merker (2007, p. 73) dice: «[El yo consciente] es único y está situado detrás del puente nasal, dentro de la cabeza. Desde allí parece que nos enfrentamos al mundo visible directamente a través de un único orificio ciclópeo de la parte frontal de la cabeza (Hering, 1879; Julesz, 1971). Sin embargo, es evidente que se trata de una mera apariencia, puesto que, si nos halláramos literalmente dentro de nuestra cabeza, al mirar deberíamos ver no el mundo, sino los tejidos anatómicos del interior de la parte frontal del cráneo. El orificio ciclópeo es una ficción neuronal conveniente por la cual “se inserta” el mundo visual distal a través de una parte ausente del cuerpo visual proximal, que está, por así decirlo, “sin cabeza” o, más precisamente,

que carece de su región facial superior (véase Harding, 1961). La somestesia, por el contrario, mantiene una continuidad ininterrumpida en toda esta región. El hueco por el que miramos el mundo delata la naturaleza simulada del cuerpo y del mundo que nos son dados en la conciencia».

[212] Ibid., p. 72. Véase Stoerig y Barth, 2001, para una simulación plausible. Esto da una cierta impresión del probable mundo sensoriomotor del niño hidranencefálico y del animal decorticado (cf. también visión ciega).

[213] White et al., 2017; Panksepp, 1998, p. 311.

[214] Merker, 2007, p. 72.

[215] Hohwy, 2013.

[216] Como le ocurrió al final, pocos años después de la operación, a consecuencia de una infección de las vías respiratorias altas (fácilmente tratable) que se detectó demasiado tarde.

[217] Véase Friston, 2005.

[218] Hohwy, 2013; Clark, 2015.

[219] Consideremos, por ejemplo, las neuronas DPM en la mosca de la fruta Drosophila. Sorprendentemente, incluso parece haber precursores primitivos en el nematodo Caenorhabditis elegans (véanse Bentley et al., 2016; Chew et al., 2018).

[220] La siguiente imagen de Freud (1925, p. 231) puede ayudarnos a imaginar la disposición funcional que acabamos de describir y, al mismo tiempo, puede permitirnos sustituir sus términos «metapsicológicos» por términos fisiológicos: «Las inervaciones catécticas son enviadas y retiradas en rápidos impulsos periódicos desde el interior hacia el sistema totalmente permeable Pcpt-Cs [el sistema de "conciencia perceptiva"]. Mientras ese sistema está catectado de esta manera, recibe las percepciones (acompañadas de conciencia) y transmite la excitación a los sistemas mnémicos inconscientes; pero cuando la catexis se sustrae, la conciencia se extingue y cesa la función del sistema. Es como si el [ello] extendiera las antenas, a través del sistema Pcpt-Cs, hacia el mundo exterior y las retirara apresuradamente tan pronto como han probado las excitaciones procedentes de él». «Catexis» es la excitación moduladora y, por ende, las «inervaciones catécticas» que palpan la percepción cortical en esta imagen figurativa son impulsos de excitación cerebral

central. En esta cita he sustituido el término de Freud «el inconsciente» por «el ello», para esquivar su fusión errónea de ambos sistemas (Solms, 2013).

[221] Merker, 2007. Panksepp y Biven (2012, pp. 404-405) utilizan el ritmo theta del hipocampo como ejemplo de en qué podría resultar esta «divisa común», fisiológicamente hablando, teniendo en cuenta que el hipocampo codifica el contexto. «En la literatura neurocientífica tradicional hay indicios de ciertos tipos de oscilaciones sincrónicas pertinentes dentro del cerebro, como los ritmos de 4-7 Hz en el hipocampo conocidos como ritmo theta, que ayudan a los animales a investigar el mundo (p. ej., el olfateo en las ratas) y, por tanto, a crear recuerdos en el hipocampo. El ritmo theta es una firma neuronal muy característica del hipocampo cuando está activo procesando información. Este ritmo resulta especialmente evidente durante la excitación artificial del sistema de BÚSQUEDA en ratas, un sistema emocional de recogida de información en toda regla, puesto que los animales olfatean e investigan su entorno (Vertes y Kocsis, 1997). En otras palabras, el ritmo de olfateo corresponde típicamente a la frecuencia en curso del theta del hipocampo. […] Esto puede poner de relieve cómo el conocimiento cognitivo emerge de las excitaciones pautadas de los procesos afectivos».

[222] Tacto es la palabra coloquial para referirse a la sensación somática, que, al igual que otras modalidades perceptivas, contiene submodalidades, como el sentido muscular y articular, el sentido de la temperatura y el sentido de la vibración.

[223] Cuando son prominentes.

[224] Solms, 2013.

[225] Podríamos decir que la acetilcolina modula la confianza en las señales de error, pero es simplificar mucho las cosas y equivaldría a hacer generalizaciones sobre la serotonina en relación con las señales de predicción, sobre la dopamina en relación con los estados activos y sobre la noradrenalina en relación con los estados sensoriales. Véase Parr y Friston, 2018, para una visión más elaborada.

[226] Física (φυσική) significa «conocimiento de la naturaleza», de toda la naturaleza, no solo de lo que se puede ver y tocar; nos ofrece la explicación más fundamental de los fenómenos naturales. Muchos dan por hecho que la física solo estudia la materia —y que, por tanto, la mente queda fuera por definición—, pero esto implicaría que la mente no forma parte de la naturaleza, lo que nos lleva precisamente

a la pregunta que da origen a este libro.

«Un fenómeno natural solo se explica físicamente en su totalidad cuando se ha remontado hasta las fuerzas últimas de la naturaleza que lo fundamentan y actúan en él» (Helmholtz, 1892). La materia resulta ser un estado energético (de ahí $E = MC^2$). Las «fuerzas últimas» explican fenómenos superficiales; no se observan directamente, se infieren. Por esta razón, se describen científicamente en términos no fenoménicos, como abstracciones.

Cf. Crítica de la razón pura de Kant: «La experiencia misma —es decir, el conocimiento empírico de las apariencias— solo es por tanto posible porque sometemos la sucesión de las apariencias, y con ella cualquier alteración, a la ley de la causalidad; y consiguientemente, las apariencias mismas, como objetos de la experiencia, solo son posibles con arreglo a esa ley».

Las abstracciones matemáticas son convencionalmente preferibles a las verbales, porque exigen que las fuerzas inferidas (y las relaciones entre ellas) sean mensurables y cuantificables. Se obtiene así una divisa común numérica por debajo de la abigarrada superficie fenoménica con la que se pueden calcular las relaciones legítimas entre las «fuerzas últimas» de la naturaleza. Como dijo Galileo: «El libro de la naturaleza está escrito en el lenguaje de las matemáticas».

[227] Llegados a este punto, a los lectores que no han estado siguiendo las notas al pie les convendría echar un vistazo al apéndice de la p. 367. Es algo técnico, pero sirve muy bien de puente hacia el capítulo siguiente.

07

El principio de la energía libre

Tenemos a un grupo de científicos en un laboratorio mirando una gran pantalla de ordenador. En la pantalla se arremolinan puntos y manchas. Podemos distinguir diferentes colores: azul, rojo, morado y algunos más. Los puntos parecen ser de diferentes tamaños, pero no se puede identificar ningún patrón en sus movimientos de enjambre. Ondulan y se intersecan como nubes de gas, ocupando desordenadamente el espacio virtual. Un reloj digital marca el paso del tiempo en segundos, y dos ejes situados en la parte inferior e izquierda de la pantalla nos dan las coordenadas de cada punto. Esto hace pensar que está ocurriendo algo mensurable, pero ¿cómo es posible cuantificar los movimientos en este caos?

Si preguntamos a uno de los físicos de la sala qué estamos viendo, responde que se trata de un proceso «estocástico», lo que no es de mucha utilidad. Estocástico quiere decir «aleatorio». Poco a poco, nos damos cuenta de que los puntos más pequeños parecen más lentos que los grandes, como si bailaran al son de una melodía ligeramente diferente. Siguiendo las instrucciones de un neurocientífico allí presente, el informático escribe algo en el teclado y la pantalla llena de remolinos se detiene de golpe. El técnico graba cuidadosamente los datos. A continuación, de nuevo a instancias del neurocientífico, teclea una ráfaga de números que cambian los valores de una serie de ecuaciones que ahora podemos ver en una pequeña pantalla junto a la más grande que estábamos mirando. Explica que está ajustando las «interacciones locales entre los subsistemas».

Vuelven a entrar por la esquina inferior izquierda de la pantalla las partículas con forma de nubes, que empiezan a arremolinarse. Esta vez, tras un breve periodo de caos, es más fácil identificar un patrón. Al principio, los puntos de colores se expanden sobre todo hacia el exterior; luego, poco a poco, se empiezan a agrupar y convergen espontáneamente en el centro de la pantalla para formar una masa amorfa. Echándole un poco de imaginación, podría tratarse de una bandada compacta de pájaros (estorninos, quizá) girando en formación. Paulatinamente, el movimiento de las partículas va restringiendo su patrón. Se empujan unas a otras como soldados que desfilan ocupando posiciones asignadas, hasta que empieza a vislumbrarse una estructura clara: cuatro capas concéntricas. En el

centro hay puntos azul oscuro rodeados por otros rojos, que a su vez están rodeados por puntos morados. Otros de color azul claro forman una especie de frontera exterior. Todo esto sucede sobre un fondo de pequeños puntos negros que parecen seguir vagando sin rumbo. El neurocientífico parece satisfecho. Pide al técnico que congele la pantalla. El reloj indica que han pasado 1.278 segundos.

Estamos en el Centro Wellcome de Neuroimagen Humana, en Queen Square (Londres), cuyo director científico es Karl Friston. Ha tenido la amabilidad de invitarnos a observar este interesante experimento: una simulación de las interacciones de corto alcance que se producen entre distintos subsistemas físicos cuando se ven sometidos a diversas fuerzas.

Las reglas que gobiernan estas partículas virtuales son del mismo carácter general que las que rigen el comportamiento de los átomos y las moléculas reales: propensiones aleatorias (pero no indiscriminadas) a atraerse y repelerse. Con toda evidencia, estas interacciones producen orden a partir del caos. Se cree que cuando la vida surgió de la sopa primigenia se produjo una ordenación espontánea semejante. Y, sin embargo, este experimento que nos ocupa se lleva a cabo en un centro de neurociencia cognitiva, justo enfrente del Hospital Nacional de Neurología y Neurocirugía. Es posible que se pregunten qué tienen que ver estas partículas virtuales con el cerebro.

Según Friston, sistemas biológicos como las células deben de haber surgido a partir de versiones complejas del mismo proceso que formó sistemas «autoorganizados» más simples, como los cristales a partir de un líquido, porque comparten un mecanismo común. Este mecanismo, comenta, es la «minimización de la energía libre» (concepto que explicaré en breve). Todos los sistemas autoorganizados, incluidos cada uno de nosotros, tienen una tarea fundamental común: seguir existiendo. Friston cree que lo logramos al reducir al mínimo nuestra energía libre. Los cristales, las células y los cerebros solo son para Friston manifestaciones cada vez más complejas de este mecanismo básico de supervivencia.[228] De hecho, en los albores mismos de la organización biológica aparecen tantos aspectos de lo que consideramos vida mental que la contribución de los cerebros reales puede empezar a parecer bastante sutil. No obstante, si nos aferramos al concepto de energía libre, todo (realmente todo) se va a aclarar.

Karl Friston es el último de los grandes científicos que han dado forma

al trabajo de mi vida. En mi opinión, es un genio y (objetivamente) el neurocientífico más influyente del mundo en la actualidad. La influencia de un científico se mide por su «índice h», que mide el impacto de sus publicaciones.[229] Como regla general, cuando el índice h es mayor que el número de años transcurridos desde el doctorado, un científico lo está haciendo bien. El índice h de Friston es de 235, el más alto de todos los neurocientíficos.[230] En un principio, alcanzó la fama gracias al «mapeo paramétrico estadístico», que permitió el análisis de neuroimágenes funcionales tan frecuente hoy en día. Sin embargo, sus trabajos sobre la «codificación predictiva» y el principio de la energía libre le dieron mucho más renombre.

A pesar de la gran reputación de Friston, durante muchos años no me interesé demasiado por su obra. Luego, en 2010, publicó un artículo con un joven psicofarmacólogo llamado Robin Carhart-Harris, a quien yo conocía un poco por su interés por el neuropsicoanálisis. Su artículo con Friston argumentaba que la concepción de Freud de la energía motriz (es decir, la «energía psíquica») era coherente con el «principio de la energía libre».[231] Como ya he explicado, Freud admitió de buena gana que era «totalmente incapaz de formarse una idea» de cómo las necesidades corporales podían convertirse en una energía mental. También escribió que esta energía podía aumentar, disminuir, desplazarse y descargarse, y que por tanto poseía todas las características de una cantidad, «aunque no tengamos medios para medirla».[232] Considerando que la intención original (abandonada posteriormente) de Freud había sido «representar los procesos psíquicos como estados cuantitativamente determinados», el artículo de Carhart-Harris y Friston fue para mí una conmoción. Si la energía mental fuera realmente isomórfica con cambios en la energía libre termodinámica, entonces, pensé, debería ser posible medirla y reducirla a leyes físicas.

Por eso me sumergí en las publicaciones anteriores de Friston y me puse en contacto con él. Nos reunimos varias veces en los años siguientes, en Londres y Fráncfort. El tema principal de nuestras conversaciones fue el papel del afecto en la vida mental. Dado que el trabajo de Friston en aquella época estaba, como el de casi todo el mundo, muy centrado en la corteza cerebral, los mecanismos de predicción que había descubierto se referían casi exclusivamente a la cognición. Por esta razón, por ejemplo, un famoso artículo suyo en el que demostraba que la codificación predictiva explica la forma en que las neuronas se comunican entre sí se titulaba «Una teoría de las respuestas corticales».[233] Debo confesar, no obstante, que nunca me había tomado el tiempo necesario para digerir a fondo algunas de sus

publicaciones más técnicas.

En 2017, nuestro congreso anual de neuropsicoanálisis (que se celebró aquel año en el antiguo University College Hospital de Londres y se centró en el tema de la codificación predictiva) invitó a Friston como ponente principal. Si no quería pasar vergüenza en mi habitual discurso de clausura, tenía que ponerme con la física. Así pues, entre muchas publicaciones de Friston, releí con atención un artículo suyo muy técnico publicado en una de las revistas de la Royal Society. Se titulaba «La vida tal y como la conocemos».[234] Con grandes esfuerzos, logré entenderlo bien por primera vez. El artículo pretendía nada menos que reducir a ecuaciones matemáticas las leyes básicas que rigen la intencionalidad.

Las implicaciones eran emocionantes. Me pareció que estas ecuaciones podrían proporcionar el avance que estaba buscando. Por ello, inmediatamente después de nuestros intercambios científicos en el congreso de 2017, le escribí sugiriéndole que pusiéramos en común nuestros conocimientos e intentáramos incorporar la conciencia al principio de la energía libre. Para mi contento, Friston estuvo de acuerdo y empezamos a colaborar en un documento que exponía lo que se convirtió en nuestro punto de vista compartido.[235]

El vínculo entre el trabajo de Friston y el mío es la homeostasis. Antes he explicado que debemos mantenernos dentro de unos límites fisiológicamente viables. Tomemos la termorregulación como ejemplo paradigmático. La temperatura corporal no admite probaturas: tenemos que mantenerla dentro del rango limitado de entre 36,5 ºC y 37,5 ºC. Con una temperatura mucho mayor morimos, y con una mucho menor… morimos. No podemos permitir que la temperatura central de nuestro cuerpo se iguale con la temperatura ambiente, como pasa con el agua caliente cuando se añade a un baño frío. El agua caliente que entra en la bañera no permanece separada de la fría en una gran burbuja bajo el grifo. En cambio, nosotros sí —es lo que debemos hacer si queremos seguir vivos—, y eso requiere trabajo. Los pacientes comatosos no pueden hacerlo, por lo que mueren de afecciones como la hipertermia (literalmente, se sobrecalientan).[236] Lo mismo ocurre con la regulación de los gases en sangre, el balance hídrico y energético, y muchos otros procesos corporales. Lo podemos aplicar incluso a las necesidades emocionales, que, como hemos visto en el capítulo 5, no son menos «biológicas» que las corporales. Permanecer dentro de los límites viables de nuestras emociones también requiere un trabajo: para mantenernos cerca de nuestros

cuidadores, para escapar de los depredadores, para librarnos de obstáculos frustrantes, *etc.* Más allá de un cierto grado de previsibilidad, el trabajo necesario para hacerlo está regulado por sensaciones.

El mecanismo que acabo de describir es una forma ampliada de homeostasis y no es nada complicado (véase fig. 12).

Un homeostato solo tiene tres componentes: un receptor (que mide la temperatura, en mi ejemplo), un centro de control (que determina cómo mantener una temperatura dentro de los límites viables: entre 36,5 ºC y 37,5 °C, en mi ejemplo) y un efector (que hace lo necesario para volver a esos límites cuando se sobrepasan). Como el mecanismo de la homeostasis es muy sencillo, puede reducirse a unas leyes físicas. De eso trataba el artículo de Friston: de las leyes básicas que rigen «la vida tal y como la conocemos».

Lo que me entusiasmó fue darme cuenta de que estas leyes —ampliadas constantemente para dar cabida a la forma menos predecible de homeostasis que subyace en el afecto— podrían explicar la conciencia: y no solo los elementos observables externos del comportamiento consciente (la fisiología objetiva del triángulo decisorio del mesencéfalo y el sistema reticular activador), sino también los elementos observados internamente: podrían explicar los sentimientos subjetivos que rigen las decisiones.

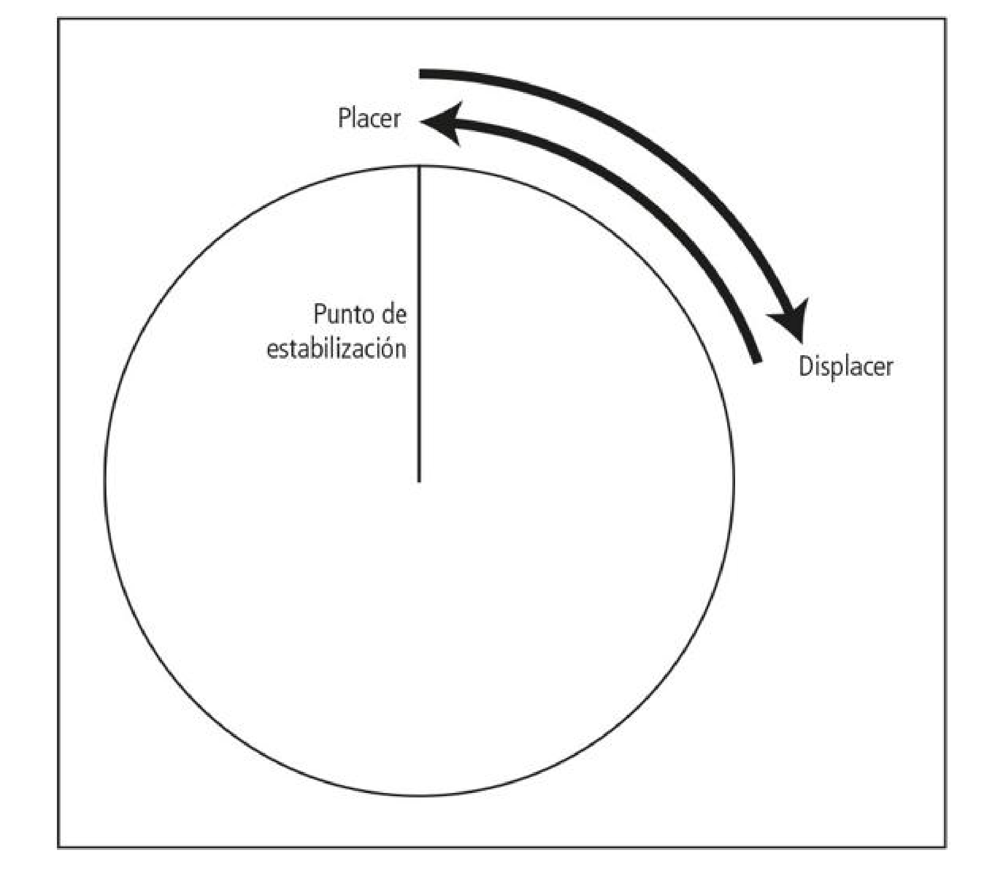


Figura 12

. Homeostasis de las sensaciones. El punto de estabilización representa los límites viables del sistema.

Aquí es donde Damasio se aparta de mí. Quizá más que ningún otro neurocientífico, Damasio había llamado la atención sobre el hecho de que el afecto (y, por lo tanto, la conciencia) es en el fondo una forma de homeostasis. Sin embargo, rechazó la inferencia adicional que yo extraje: si la conciencia es homeostática, y si la homeostasis se puede reducir a leyes físicas, lo mismo ocurre con los fenómenos de la conciencia.[237] El mecanismo de la conciencia, como el mecanismo del movimiento de los cuerpos celestes y todo lo demás en la naturaleza, debe poderse explicar a través de las leyes y, por lo tanto, ser predecible de algún modo, aunque solo sea de forma probabilística. Cuando Damasio leyó un borrador del artículo que escribí con Friston, mantuvimos una angustiosa conversación en su

despacho. No entendía por qué yo intentaba reducir la conciencia a lo que él llamaba «algoritmos». Se trata del caso habitual de un científico que se opone a las implicaciones derivadas de una idea que él mismo ha aportado al mundo, como en la respuesta de Einstein a la mecánica cuántica: «Dios no juega a los dados con el universo».[238]

A Damasio le planteé brevemente el argumento que voy a explicarles ahora.[239] Para hacer las cosas bien, primero debo presentar la física que tuve que aprender para entender el principio de la energía libre. Respiren hondo.

La esencia de la homeostasis es que los organismos vivos deben ocupar una gama limitada de estados físicos: sus estados viables, o estados valorados o preferidos, o lo que Friston llama (refiriéndose a todo lo anterior), sus estados «previsibles». No podemos permitirnos una dispersión por todos los estados posibles. Este imperativo biológico tiene un profundo vínculo con uno de los conceptos explicativos más básicos de la física, a saber, la entropía. La mayoría de la gente tiene una comprensión intuitiva de lo que es la entropía. Piensan que es una tendencia natural al desorden, la disipación, la disolución y cosas por el estilo. Las leyes de la entropía son las que hacen que el hielo se derrita, las baterías se descarguen, las bolas de billar se detengan y el agua caliente se mezcle con la fría.

La homeostasis funciona en sentido contrario. Combate la entropía. Garantiza que ocupemos un rango de estados limitado. Así es como nos mantiene la temperatura necesaria y como nos mantiene vivos: como evita que nos disipemos. Los seres vivos deben combatir uno de los principios fundamentales de la física: la segunda ley de la termodinámica.

La primera ley de la termodinámica se refiere a la conservación de la energía.[240] Establece que la energía no puede crearse ni destruirse; solo se puede convertir de un tipo de energía a otro y fluir de un lugar a otro. (También sabemos, gracias a Einstein, que puede convertirse en materia).

La segunda ley establece que los procesos naturales son siempre irreversibles.[241] Por lo tanto, el agua del baño caliente que se mezcla con la fría no puede volver a separarse. Del mismo modo, la energía del calor no puede volver al carbón quemado que lo produjo y la energía que se desperdició en el proceso no puede volver a él. Esto se debe a la entropía, que, en consecuencia, siempre aumenta a gran

escala.[242] Puede incluso que la entropía sea la base física del hecho de que el propio tiempo parezca tener una dirección y un flujo.

En termodinámica, hay dos condiciones de energía: útil e inútil. La «utilidad» de la energía se define por su capacidad para realizar un trabajo. Por ejemplo, la energía de un trozo de carbón se puede quemar para producir calor,[243] que puede hervir agua para producir vapor, que puede mover un motor, pero en cada paso de este proceso se perderá algo de energía. Es decir, nunca se puede emplear útilmente toda ella.

Combinando estos hechos, aprendemos lo siguiente: a medida que se agota la energía útil de un sistema, aumenta su entropía. Esto quiere decir que la capacidad del sistema para llevar a cabo un trabajo siempre disminuye. Por tanto, la entropía está asociada a la pérdida de energía útil, porque la energía ya no está disponible para llevar a cabo ese trabajo. La segunda ley es una declaración del hecho ineludible de que durante cualquier proceso natural se perderá parte de la energía para el trabajo útil.[244]

Hace un momento he mencionado algunas interpretaciones intuitivas de la «entropía», pero la definición técnica y formal de la entropía en física alude al número de estados diferentes que puede ocupar un sistema dado.[245] La entropía viene determinada por el número de estados microscópicos posibles que darían lugar al mismo estado macroscópico. En pocas palabras: cuanto menor sea el número de estados posibles, menor será la entropía.

La homeostasis pone límite a la gama de estados macroscópicos que podemos ocupar sistemas como ustedes o yo. Recordemos que la homeostasis nos mantiene vivos mediante la realización de trabajo efectivo; por lo tanto, si la entropía supone una pérdida de la capacidad de trabajo, es «mala» desde nuestro punto de vista en cuanto que sistemas biológicos. La función más básica de los seres vivos es combatir la entropía.

El ejemplo que suelen utilizar los físicos para ilustrar estos conceptos consiste en trasvasar un gas comprimido a una cámara vacía de mayor tamaño a través de una pequeña abertura. A medida que las moléculas se mueven aleatoriamente, exploran la cámara y se expanden para ocupar todo el espacio disponible. Cuanto más tiempo pase, habrá más puntos en los que se puede encontrar una molécula. La única forma de revertir este proceso —es decir, la única forma de devolver el gas a su

recipiente original— es mediante trabajo.

Pensemos en la entropía en términos del número de ubicaciones en las que puede encontrarse cada molécula en un momento dado, lo que en realidad es una afirmación acerca de la probabilidad. La probabilidad estadística de que cada molécula ocupe una posición específica disminuye a medida que aumenta la entropía: a medida que el gas se expande, la posición de cada molécula se vuelve menos predecible. Aumentar la entropía significa disminuir la predictibilidad.

Este aspecto es importante porque, a diferencia de las demás leyes de la termodinámica, las leyes de la probabilidad se aplican a todo, no solo a las cosas materiales. Al igual que la entropía de un gas en una cámara puede definirse de forma probabilística, lo mismo ocurre con la entropía asociada a un proceso psicológico de toma de decisiones. En ambos casos, la entropía aumenta con la aleatoriedad de los resultados posibles. La «entropía» asociada a los gases en expansión y a las opciones de expansión es lo mismo. No todo lo que existe en la naturaleza son cosas visibles y tangibles, pero todo se somete por igual a las leyes de la probabilidad. De ahí que la probabilidad afecte al corazón de la física moderna, donde la materia ya no se considera un concepto fundamental y las partículas clásicas han desaparecido.[246]

En mi ejemplo físico de entropía, cuando el gas se introducía a presión en su primer recipiente y sus moléculas se empaquetaban muy juntas, hacían falta menos bits de información para describir la ubicación real de cada molécula que los que se necesitaban una vez que se liberaba el gas para que llenara todo el espacio disponible en la cámara grande. En ciencia de la información, un «dígito binario» (lo que normalmente conocemos como «bit») es la unidad básica de información. Un bit puede tomar uno de dos valores opuestos, por ejemplo, sí frente a no, encendido frente a apagado, positivo frente a negativo. Estos estados suelen representarse como uno frente a cero.[247]

Para describir el gas al principio se necesitan menos bits de información, pero a medida que se expande, aumenta el número de estados posibles que puede ocupar cada molécula. Así pues, la entropía (que físicamente se suele medir en términos de intercambio de calor junto con temperatura) también puede medirse en bits: cuanta más información sea necesaria para describir el microestado de un sistema (es decir, el estado de todas y cada una de las moléculas), mayor será la entropía termodinámica. O dicho de forma todavía más sencilla: cuantas más preguntas sí/no haya que responder para

describir un sistema, mayor será su entropía. Así, cuando los microestados de un sistema tienen valores de probabilidad bajos, cada medición del sistema transmite más información que si tuvieran una probabilidad alta, porque habría que responder a más preguntas binarias para describir el estado del sistema en su totalidad.

La entropía es mínima cuando la respuesta a cada pregunta sí/no es totalmente predecible, es decir, cuando no se aprende nada y no se gana información. El contenido informativo de una moneda lanzada a cara o cruz es de un bit, porque las probabilidades de que salga cara son del 50 por ciento, mientras que se obtiene información cero lanzando una moneda con dos caras, porque las probabilidades de que salga cara son del 100 por ciento. No se aporta ninguna información porque la respuesta es totalmente previsible. La entropía mide la cantidad media de información que se obtiene tras múltiples mediciones de un sistema. Así, la entropía de una serie de mediciones es su información media, su incertidumbre media.

Para hacernos una idea de la importancia que tiene esto para la comprensión neurocientífica de la conciencia, debemos recordar que los patrones sincronizados de ondas lentas en un electroencefalograma son más ordenados (más predecibles) que los rápidos desincronizados (erráticos). Por lo tanto, los patrones de «baja excitación» contienen menos información que los de «alta excitación» (véase fig. 10). Los de alta excitación contienen más incertidumbre.[248] Así, los valores de entropía del electroencefalograma son más altos en los pacientes mínimamente conscientes que en los vegetativos.[249] Tiene sentido: la actividad cortical en el encéfalo consciente comunica más información que durante el sueño profundo. Sin embargo, ocurre algo extraño: si más información significa más incertidumbre y, por tanto, más entropía, entonces (dado que los seres vivos deben combatir la entropía) la actividad durante la vigilia es menos deseable, biológicamente hablando, que el sueño profundo.[250] Sé que es una idea contraintuitiva, pero será más comprensible a medida que avancemos.[251]

La relación entre entropía e información se formalizó en una famosa ecuación del ingeniero eléctrico y matemático Claude Shannon. A través de este avance, Shannon incorporó él solo la «información» a la física, que desde entonces lo ha adoptado como uno de sus conceptos básicos, especialmente en la mecánica cuántica.[252] Sobre la base del trabajo de Shannon, el físico Edwin Thompson Jaynes argumentó que la entropía termodinámica debería considerarse una aplicación de la entropía de la información.[253] Por lo tanto, la definición de Shannon es más fundamental que la termodinámica: una definición

abstracta de la entropía en términos de dinámica de la información tiene más aplicación en general que una concreta en términos de dinámica del calor. Por tanto, las leyes de la termodinámica podrían considerarse un caso especial de las leyes más profundas de la probabilidad. Este detalle es importante, porque las leyes termodinámicas solo se aplican a sistemas materiales (tangibles, visibles) como los cerebros, mientras que las leyes de la información se aplican también a sistemas inmateriales (intangibles, invisibles) como las mentes.

No obstante, probabilidad no es lo mismo que información. La información, en el sentido que le da Shannon, conlleva el factor adicional de la comunicación. De ahí el título de su influyente artículo que fundó la ciencia de la información: «Una teoría matemática de la comunicación».[254] A diferencia de las probabilidades, que existen por sí mismas, la comunicación requiere un emisor y un receptor de la información. (El comunicador no tiene por qué ser una persona. Puede ser un libro, por ejemplo, o cualquier sistema que tenga una información de la que un receptor pueda aprender).

Esto plantea grandes problemas para cualquier suposición teórica de que la conciencia solo es información.[255] La pregunta que se plantea es: ¿cuál es el emisor y cuál es el receptor de la información, integrada o no? Por eso no estoy satisfecho con los modelos de flujo de información utilizados por los científicos cognitivos, porque eluden la pregunta de dónde está el sujeto, el receptor. De este modo, parafraseando a Oliver Sacks, la psique queda excluida de la ciencia cognitiva.

Sin embargo, esta omisión del sujeto que experimenta plantea una cuestión más importante aún, quizá la más relevante a la que se enfrenta la ciencia cognitiva hoy en día: sin un observador, ¿cómo y por qué se produce el procesamiento de la información (es decir, la formulación de preguntas y la respuesta a las mismas)?

El descubrimiento de Shannon de la información como entropía llevó al físico John Wheeler a proponer una interpretación «participativa» del universo.[256] Según Wheeler, las cosas solo surgen en la forma en que lo hacen (es decir, como fenómenos observables) en respuesta a las preguntas que formulamos. Los fenómenos, como tales, solo existen a los ojos de quien los contempla, de un observador participante, de quien formula preguntas. Repitiendo la célebre frase de Wheeler, «eso surge de los bits» (cuando nos referimos a cosas observables).[257] Por tanto, la información es «física» no solo porque está involucrada en las leyes de la física, sino también porque es la

base de todos los fenómenos observables. Así es como las fuerzas y energías abstractas se hacen observables y mensurables: «Lo que llamamos realidad surge en última instancia del planteamiento de preguntas sí/no y del registro de las respuestas suscitadas por el equipo; en resumen, […] todas las cosas físicas tienen que ver en un principio con la teoría de la información».[258]

Las modalidades sensoriales del sistema nervioso generan «respuestas suscitadas por el equipo» a las preguntas que hacemos al universo. Las respuestas sensoriales dan lugar a los fenómenos (las «cosas») que experimentamos. Por tanto, la propia experiencia surge de la comunicación entre un receptor de información (un observador participante) y un emisor de información; entre alguien que pregunta y las respuestas que registra. Con todo, todavía nos queda una pregunta: ¿de dónde vienen los que preguntan?

Antes de adentrarme en mayores complejidades, hagamos una pausa para hacer balance. He transmitido tres aspectos importantes. El primero es que la información media de un sistema es la entropía de ese sistema (es decir, la entropía de un sistema es una medida de la cantidad de información necesaria para describir su estado físico). El segundo es que los sistemas vivos deben combatir la entropía. Estos dos elementos implican conjuntamente que debemos minimizar la información que procesamos (y aquí me refiero a la información en el sentido que le da Shannon, por supuesto; es decir, debemos minimizar nuestra incertidumbre). Todo lo demás que voy a decir en este capítulo, y en los dos siguientes, se desprende de esta sencilla pero sorprendente conclusión.

Y esto nos lleva al tercer aspecto importante que hemos aprendido hasta ahora: los sistemas vivos combatimos la entropía mediante el mecanismo de la homeostasis. En resumen: recibimos información sobre nuestra posible supervivencia haciéndonos preguntas (es decir, tomando medidas) sobre nuestro estado biológico en relación con el desarrollo de los acontecimientos. Cuanto más inciertas sean las respuestas (es decir, cuanta más información contengan), peor para nosotros, pues significa que estamos incumpliendo nuestra obligación homeostática de ocupar estados limitados (nuestros estados previsibles).

La naturaleza de las preguntas que nos hacemos viene determinada en parte por nuestra especie. Los tiburones pueden respirar bajo el agua; los humanos, no. Así que tenemos necesidades distintas y esperamos ocupar estados distintos. Estas necesidades vienen determinadas por la selección natural. La homeostasis consiste en mantenerse dentro del

nicho evolutivo propio según una concepción amplia del mismo. Por eso, cada especie debe plantearse preguntas como: ¿soy capaz de respirar aquí? Nuestra supervivencia depende de las respuestas que recibamos.

Por cierto, ¿por qué deberíamos pensar en las necesidades biológicas como expectativas? Este lenguaje puede sorprender aquí, pero expresa una continuidad profunda que resultará importante más adelante. Si sirve de ayuda, podemos intentar adaptar la perspectiva de la propia evolución, en lugar de la de una criatura individual. La selección natural adaptó cada especie a su nicho ecológico; la supervivencia de cada criatura depende solo de cosas que se encuentran real y fiablemente en su hábitat natural. Así pues, necesitamos aire porque tenemos la previsión de tenerlo.

Ahora puedo volver a la pregunta profunda que planteaba antes: ¿de dónde vienen los observadores participantes? En otras palabras: ¿cómo y por qué, en términos físicos, surge la formación de preguntas?[259]

He aquí una breve historia de la idea de autoorganización. El primero en utilizar el término fue Immanuel Kant, en Crítica del juicio, en 1790. Kant alegaba que los seres vivos tienen «objetivos» y «propósitos» intrínsecos, lo que, para él, solo puede ser cierto si sus mecanismos constitutivos son simultáneamente un medio y un fin. Estas entidades «teleológicas» (es decir, entidades con objetivos y propósitos intrínsecos), decía Kant, deben comportarse intencionadamente. «Solo de esta forma y en estos términos puede un producto ser organizado y autoorganizado y, como tal, ser considerado un fin físico». Kant creía que la ciencia no podía explicar cómo surgirían estos seres: nunca podría haber «un Newton que haga concebible la producción de una brizna de hierba».

Entonces, Darwin descubrió la selección natural. Como ahora sabemos, la selección natural da lugar a los objetivos y propósitos intrínsecos de supervivencia y reproducción. Ambas cosas resultan ser manifestaciones de la autoorganización.[260] Con la visión de Darwin, la cuestión del origen y la composición de los seres teleológicos pasó a estar al alcance de la ciencia.[261] Solo quedaba concretar los detalles.

A mediados del siglo XX, cuando Norbert Wiener, el matemático que fundó la «cibernética», añadió la noción de retroalimentación al

concepto de comprensión de la información de Shannon, se dio un paso importante. Según Wiener, un sistema puede alcanzar su objetivo (su «estado de referencia») recibiendo retroalimentación sobre las consecuencias de sus acciones. La retroalimentación incluye señales de error, que miden las desviaciones del estado de referencia y que se pueden utilizar para ajustar las acciones del sistema y mantenerlo en marcha. Así pues, la homeostasis resulta ser un caso específico de un principio cibernético más general: es una especie de retroalimentación negativa.

William Ross Ashby utilizó esta noción de retroalimentación, combinada con la física estadística presentada anteriormente, para revelar cómo se desarrolla de forma natural la autoorganización.[262] Ashby demostró que muchos sistemas dinámicos complejos evolucionan automáticamente hacia un punto de estabilización que describió como un «atractor» en una «cuenca» de estados circundantes. La evolución posterior de tales sistemas tiende a ocupar estados limitados.

Espero que esta tendencia a ocupar estados limitados les resulte ya familiar: no es sino una tendencia a combatir la entropía. Según Friston, esta tendencia desencadena formas de autoorganización cada vez más elaboradas. Una vez creada la posibilidad para ello entre los subsistemas de su sopa primigenia simulada descrita al principio de este capítulo, observó que su comportamiento se desarrollaba en tres etapas:

1) con ciertos parámetros de corto alcance, los puntos saltaban por todas partes;

2) con otros parámetros se unían en estructuras cristalinas estables;

3) con otros parámetros distintos mostraban comportamientos más complejos: tras unirse, se agitaban inquietos unos contra otros, adoptando posiciones específicas dentro de una estructura dinámica.

Así es como Friston lo describió con sus propias palabras (ignoren el lenguaje técnico, solo quiero dar una impresión visual de lo que vio):

Estos comportamientos van de un comportamiento gaseoso (en el que

los subsistemas se acercan ocasionalmente lo bastante como para interactuar) hasta un hervidero de actividad, cuando los subsistemas se ven obligados a unirse en el fondo del pozo de potencial. En este régimen, los subsistemas se acercan lo suficiente como para que la ley del cuadrado inverso los haga estallar, lo que hace pensar en las colisiones de partículas subatómicas en física nuclear. Con parámetros de valores determinados, estos sucesos esporádicos y críticos pueden hacer que la dinámica no sea ergódica, con fluctuaciones imprevisibles de gran amplitud que no se estabilizan. En otros regímenes, emergen estructuras más cristalinas con interacciones atenuadas y baja entropía estructural (configuracional). Sin embargo, para la mayoría de los valores de los parámetros, el comportamiento ergódico emerge a medida que el conjunto se aproxima a su atractor global aleatorio (normalmente a partir de 1.000 s): en general, los subsistemas se repelen entre sí inicialmente (de forma muy parecida a las ilustraciones del big bang) y luego retroceden hacia el centro, encontrándose a medida que se fusionan. Las interacciones locales median entonces en una reorganización en la que los subsistemas se desplazan (a veces hacia la periferia) hasta que los vecinos se empujan sutilmente unos a otros. Desde el punto de vista de la dinámica, la sincronización transitoria puede aparecer como ondas de estallido dinámico. […] En resumen, el movimiento y la dinámica electroquímica se parecen mucho a una sopa agitada (no muy diferente de las erupciones en la superficie del Sol), pero ¿tiene algún tipo de autoorganización más allá de esto?[263]

La respuesta a su última pregunta resulta ser afirmativa. Emerge una estructura dinámica compleja, en la que los subsistemas densos, tras separarse de su entorno, forman estructuras en capas concéntricas, cada una de las cuales tiene un núcleo interior y una superficie exterior, que a su vez se divide en dos subcapas (véase fig. 13). Las subcapas de la superficie particionada muestran patrones muy interesantes de interacción tanto con el núcleo interno como con el entorno circundante. Los estados de la subcapa externa están influidos por los del entorno externo, y estos, a su vez, influyen en los de los subsistemas internos, pero esta influencia no es recíproca (en otras palabras, los constituyentes internos del núcleo no tienen ningún impacto en la subcapa externa). Del mismo modo, los estados de la subcapa interna se ven afectados causalmente por los del núcleo interno, y estos, a su vez, influyen en los del entorno externo, pero la línea de influencia tampoco es recíproca. Esta disposición de dependencias causales define las propiedades de lo que se conoce como una manta de Markov.[264]

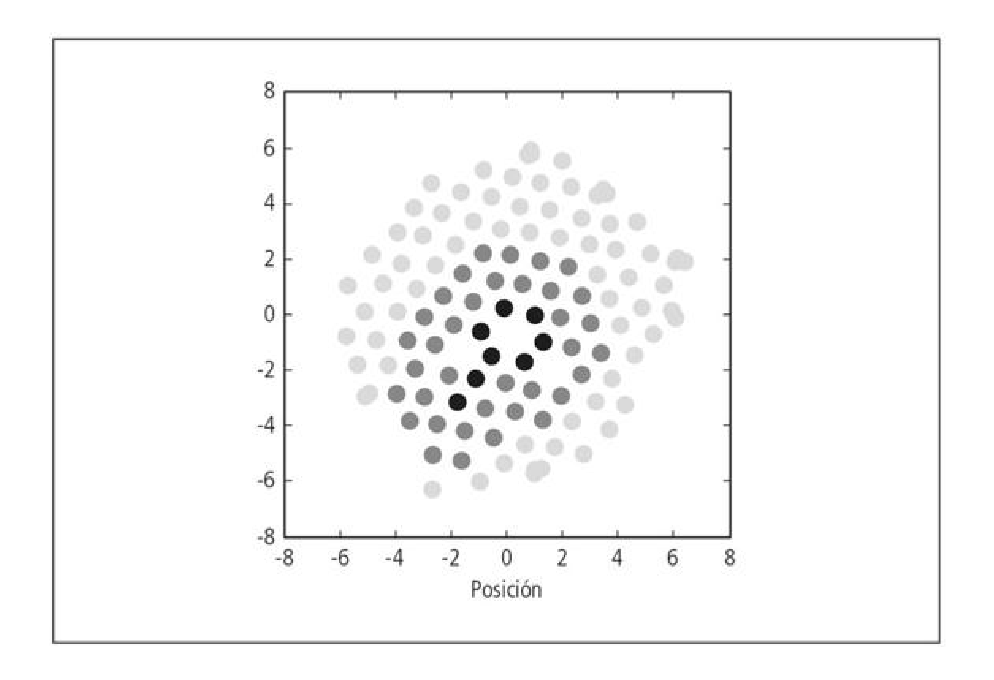


Figura 13

. Un sistema autoorganizado con su manta de Markov. En esta imagen, el núcleo interno del sistema está representado por los puntos negros y las capas que lo rodean, por los puntos gris oscuro: la manta. Los puntos de color gris claro son externos al sistema. (La imagen original de Friston diferenciaba las subcapas de la manta. Para que conste, ya que sus colores se mencionan en el texto: los puntos externos eran azul pálido, los internos azul oscuro, los sensoriales morados y los activos rojos).

La manta de Markov es un concepto estadístico que separa dos conjuntos de estados entre sí. Estas formaciones presentan una partición de los estados en internos y externos, es decir, en un sistema y un no sistema, de modo tal que los estados internos quedan aislados de los externos al sistema. Es decir, los estados externos solo pueden ser «percibidos» indirectamente por los internos como estados de la manta. Además, una manta de Markov se divide a su vez en subconjuntos que dependen causalmente de los estados del conjunto externo y subconjuntos que no dependen causalmente (directamente) de los estados del conjunto externo. Estos estados de la manta se denominan «sensoriales» y «activos», respectivamente.

La formación de una manta de Markov divide así los estados de un sistema en cuatro tipos: internos, activos, sensoriales y externos. En este esquema, los estados externos no forman parte de la entidad autoorganizada. De manera significativa, las dependencias entre estos cuatro tipos de estado crean una causalidad circular. Los estados externos influyen sobre los internos a través de los estados sensoriales de la manta, mientras que los estados internos se acoplan de nuevo a los externos a través de sus estados activos. De este modo, los estados internos y externos se causan mutuamente en un esquema circular. Dicho de otro modo, los estados sensoriales retroalimentan las consecuencias del efecto sobre los estados externos de los estados activos, ajustando así las acciones posteriores del sistema.

No es casual que todo esto nos haga pensar en el ciclo percepción-acción de los organismos vivos. También debemos recordar lo que he dicho en el capítulo anterior sobre la causalidad circular entre el sistema reticular activador, el prosencéfalo y el triángulo decisorio del mesencéfalo. Este es precisamente el valor de estos modelos abstractos. Revelan formalismos profundamente regulares que pueden reconocerse en una amplia gama de sustratos que nos permiten comprender de forma diferente la estructura de los seres vivos.

Si miramos con atención, veremos mantas de Markov por todas partes. Una membrana celular tiene las propiedades de una manta de Markov, al igual que la piel y el sistema musculoesquelético del cuerpo en su conjunto (que a su vez está compuesto por células). Lo mismo ocurre con todos los orgánulos, órganos y sistemas fisiológicos. El encéfalo (en realidad todo el sistema nervioso), que regula los demás sistemas del cuerpo, tiene, por tanto, una manta de Markov. De hecho, se trata de una metamanta, ya que envuelve a todas las demás mantas. Los sistemas autoorganizados siempre pueden estar compuestos por sistemas autoorganizados más pequeños, no siempre hasta el final, pero sí a lo largo de un camino vertiginosamente largo. Este es el tejido básico de la vida, miles de millones de pequeños homeostatos envueltos en sus mantas de Markov.

Nos acercamos a la respuesta a la pregunta que planteaba antes: ¿por qué y cómo surge, en términos físicos, la formulación de preguntas? Todavía no hemos llegado a puerto, pero, por lo que he contado hasta ahora, es razonable pensar que la auténtica mismidad de un sistema dinámico complejo está constituida por su manta. Estos sistemas autoorganizados surgen al distanciarse de todo lo demás. Una vez separados, solo pueden registrar sus propios estados: el mundo que no pertenece al sistema solo puede «conocerse» de forma indirecta, a través de los estados sensoriales de la manta del sistema. Voy a

postular que estas propiedades de la autoorganización son, de hecho, las condiciones previas esenciales de la subjetividad.

Pero antes quiero hacer una distinción: lo que proporciona la base elemental de la mismidad es el instinto de conservación de estos sistemas, su tendencia a independizarse de su entorno y luego mantener activamente su propia existencia. Y es la naturaleza aislada de tales sistemas, el hecho de que solo puedan registrar lo que no son ellos a través de los estados sensoriales de sus propias mantas, lo que constituye la base elemental de la subjetividad: el «punto de vista» de un yo secuestrado.

Por supuesto, esto no implica que todo sistema autoorganizado tenga una subjetividad sintiente. Todavía estamos lejos de poder identificar las propiedades específicas que debe presentar un sistema autoorganizado para poder ser consciente. Sin embargo, incluso sin que la conciencia tenga que intervenir, parece que hemos dado con un prototipo físico para el problema de las otras mentes. La propia naturaleza de una manta de Markov hace que genere una división de los estados en «sistema» y «no sistema», de forma que los estados que no están en el sistema se ocultan en el interior del sistema, y viceversa.[265]

Volvamos a la sopa primigenia de Friston, donde las cosas se ponen todavía más extrañas. Tras generar espontáneamente un complejo sistema dinámico autoorganizado, Friston comprobó si este conjunto permitía predecir estados externos a partir de los estados internos del sistema. De ser así, según Friston, podría ser que los estados internos de un sistema hayan modelado sus estados externos a lo largo del tiempo. Y también podría decirse que representan esos sucesos externos dentro de sí mismos. Sé que suena a magia, pero eso simplemente quiere decir que el sistema se ha ajustado a los patrones de los sucesos externos, que se ha acomodado a ellos. (Para simplificar todavía más: por eso se puede predecir la dirección típica del viento en una zona a partir de la inclinación de los árboles cuando no sopla el viento. La inclinación de los árboles «representa» la dirección típica del viento porque han crecido con ese ángulo para adaptarse a él).

Friston examinó el estado funcional de los subsistemas internos de su organismo simulado y lo que encontró fue justamente esta capacidad predictiva.

La dinámica interna que predice [un suceso externo] parece emerger

en sus fluctuaciones antes del propio suceso, como cabría esperar si los sucesos internos estuvieran modelando los sucesos externos. Curiosamente, el subsistema del que había mejores predicciones era el más alejado de los estados internos. Este ejemplo ilustra cómo los estados internos infieren o registran sucesos distantes de un modo similar a la percepción de sucesos auditivos a través de ondas sonoras, o al modo en que los peces perciben el movimiento en su entorno. [Los] subsistemas cuyo movimiento podría predecirse con fiabilidad [son] los más significativos en la periferia del conjunto, donde la libertad de movimientos es mayor. Estos movimientos se acoplan a los estados internos —a través de la manta de Markov— mediante una sincronía generalizada.[266]

Tras observar esta sincronía, por la que los estados internos del sistema modelaban sucesos físicamente distantes, Friston concluyó que los estados internos son capaces de «inferencias».

Esta resulta ser la propiedad más significativa de tales sistemas. La manta de Markov dota a los estados internos de los sistemas autoorganizados de la capacidad de representar probabilísticamente los estados externos ocultos, de modo que el sistema puede inferir las causas ocultas de sus propios estados sensoriales, algo parecido a la forma en que funciona la percepción. Esta capacidad, a su vez, le permite actuar intencionadamente sobre el medio externo, basándose en sus estados internos, con acciones que se asemejan a la actividad motriz.

De este modo, el sistema se mantiene y se renueva frente a las perturbaciones externas.[267] El mero hecho de ser un sistema autoorganizado basta para conferirles una finalidad a él y a cada una de sus partes, y esa es la función de los estados activos de la manta: manipular el entorno para mantener la integridad del sistema. Lo que significa que, junto con un yo cerrado, un punto de vista subjetivo, una finalidad y la capacidad de sentir y de actuar, la mera presencia de una manta de Markov conlleva algo parecido a la agencia.[268] Puede que no parezca especialmente dominante en la forma en que aparece en las simulaciones de Friston: los puntos azul claro (externos) influyen en los azul oscuro (internos) a través de los morados (sensoriales), mientras que los puntos azul oscuro se acoplan a los azul claro a través de los rojos (activos). No obstante, espero que se entienda por qué los sistemas biológicos autoorganizados deben inferir las causas ocultas de sus estados sensoriales, aunque por ahora no sea de forma consciente. Si no lo hicieran, dejarían de existir. Están

obligados a modelar dependencias causales en el mundo, de manera que sus acciones en ese mundo garanticen su supervivencia.

De ahí procede el concepto de «estados previsibles» y la razón por la que los sistemas biológicos autoorganizados son homeostáticos. La homeostasis parece haber surgido con la autoorganización. Los estados sensoriales y activos de una manta de Markov son simplemente los receptores y efectores de un sistema autoorganizado, y el modelo de estados externos que genera es su centro de control.

Los sistemas biológicos autoorganizados deben poner a prueba sus modelos del mundo, y si el mundo no les devuelve las respuestas esperadas, deben hacer urgentemente algo diferente o morirán. Las desviaciones de los estados previsibles son, por tanto, una forma fundamental de las «respuestas suscitadas por el equipo» de Wheeler. Así es como surge la formulación de preguntas: la autoorganización llama a la existencia a los observadores participantes. La pregunta que siempre se hace un sistema autoorganizado es simplemente: ¿sobreviviré si hago eso? Cuanto más incierta sea la respuesta, peor para el sistema.

La relación entre los estados activos de una manta de Markov y la supervivencia, a través de la homeostasis, se ilustra mejor mostrando lo que ocurrió cuando Friston dañó la manta de los sistemas autoorganizados en sus simulaciones experimentales. Lo hizo impidiendo selectivamente que los estados sensoriales de la manta influyeran en sus estados activos (véase fig. 14, gráficos b, c y d). En ausencia de los estados activos habituales de la manta de Markov, se produce un caos entrópico y el sistema se disipa rápidamente. Es decir, deja de existir.

Lo que acabo de ilustrar en términos formales (utilizando conceptos de la física estadística) es la estructura básica de cualquier sistema dinámico autoorganizado. Friston resume las relaciones lícitas del siguiente modo: «un sistema ergódico aleatorio que cuente con una manta de Markov parecerá que mantiene activamente su actividad estructural y dinámica».[269] (Un sistema «ergódico» ocupa estados limitados). Este tipo de actividad cumple los criterios de Kant expuestos anteriormente: una manta de Markov es tanto el fin como el medio por el que un sistema autoorganizado persiste en el tiempo, algo que ocurre de forma natural. De esta forma, estos sistemas parecen tener mente propia, aunque muy primitiva (y no consciente).

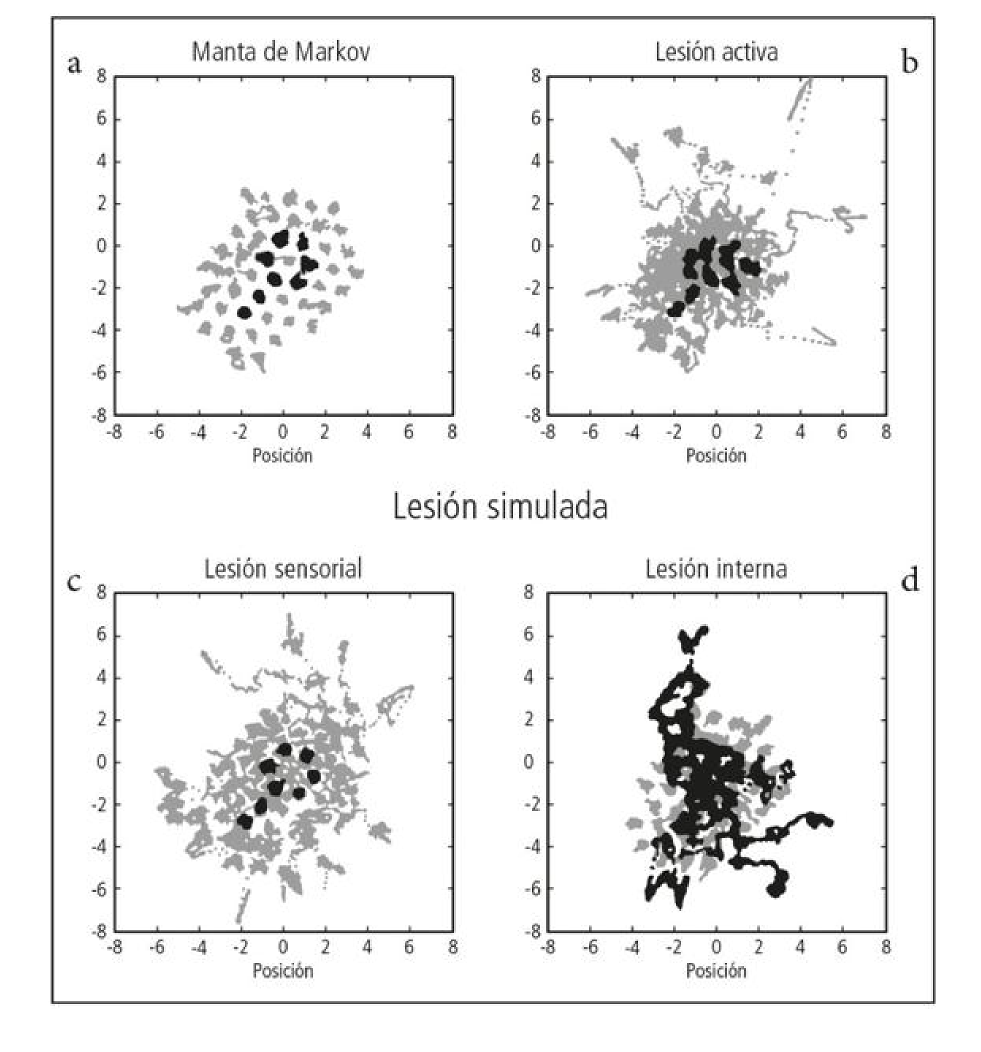


Figura 14

. Efectos entrópicos producidos al dañar ligeramente la manta de Markov de un sistema autoorganizado.

Si este argumento es correcto, entonces debería ser posible identificar la aparición de la autoorganización en cualquier conjunto arbitrario de subsistemas acoplados con interacciones de corto alcance. Esto es precisamente lo que demostró el experimento de Friston de la sopa primigenia. A partir de su experimento, abstrajo cuatro propiedades fundamentales de todos los sistemas biológicos autoorganizados:

1) son ergódicos;

2) están equipados con una manta de Markov;

3) muestran una inferencia activa;

4) tienen instinto de supervivencia.

¿Qué es exactamente el principio de la energía libre que da título a este capítulo? Para empezar, es algo muy difícil de explicar. Esto se debe en gran medida a las ecuaciones de Friston, que, como descubrí cuando me obligué a entender «la vida tal y como la conocemos», son tan abstrusas como opacas. Por lo tanto, voy a intentar explicarles el principio solo con palabras. Las ecuaciones siempre pueden traducirse en palabras porque no son más que enunciados de relaciones. Dicho esto, dado que la ecuación básica explica nuestro objetivo y propósito central en la vida, así como el de todo aquello que se ha vivido, probablemente sea interesante verla en su notación canónica al menos una vez:

$$A = U - TS$$

donde A es la energía libre, U es la energía interna total, T es la temperatura y S es la entropía.

¿Qué quiere decir esto? En termodinámica, la energía libre de un sistema es igual a la cantidad total de energía contenida en el sistema menos la parte de esa energía que se está empleando en el trabajo efectivo y, por lo tanto, no es libre.[270] Así pues, la ecuación solo dice: «La energía libre es igual a la energía interna total menos la energía ya utilizada». ¿Qué puede haber más sencillo? La energía libre es lo que queda cuando sacamos de la ecuación la energía que no es libre (es decir, cuando sacamos la «energía ligada»).[271]

La ecuación que acabo de describir cuantifica con precisión la energía libre en contextos termodinámicos básicos, pero no en contextos químicos, donde hay que dejar margen para las moléculas adicionales que se forman en algunos procesos a distintas temperaturas y presiones. Por ello, los químicos utilizan una versión ligeramente diferente de la ecuación.[272] Para diferenciarlas, la energía libre de

tipo termodinámico clásico se conoce como «energía libre de Helmholtz» (por Hermann von Helmholtz, miembro destacado de la Sociedad de Física de Berlín), y la de tipo químico-ensamblado se conoce como «energía libre de Gibbs».

Friston utiliza una tercera versión de la misma ecuación para cuantificar la energía libre en contextos de información. Llamaré a este tipo de energía libre «energía libre de Friston». La ecuación pertinente dice que «la energía libre de Friston es igual a la energía media menos la entropía».[273] Aquí, la «energía media» se refiere a la probabilidad previsible de que ocurra un suceso de acuerdo con un modelo, y «entropía» se refiere a la incidencia real de que ocurra. Así pues, la energía libre de Friston es la diferencia entre la cantidad de información que se espera obtener de una muestra de datos (de una secuencia de sucesos) y la cantidad de información que realmente se obtiene de ella. (Recordemos que la entropía de un sistema predictivo es su información media, donde el aumento de la información significa la disminución de la probabilidad).[274] La ecuación «la energía libre de Friston es igual a la energía media menos la entropía» dice básicamente lo mismo que la ecuación «la energía libre de Helmholtz es igual a la energía interna total menos la energía que no está disponible para el trabajo».[275] Esto se debe a que la energía libre de Friston es similar a la energía libre de Helmholtz, en la que hay un intercambio de información en oposición a un intercambio termodinámico entre el sistema y su entorno.[276]

Si los sistemas biológicos deben minimizar su entropía y la entropía es información media, cabe deducir que deben mantener al mínimo el flujo de información que procesan. Deben minimizar los imprevistos. Esto se conoce técnicamente como «sorpresa». Al igual que la entropía, la sorpresa es una función decreciente de la probabilidad: a medida que la probabilidad disminuye, la sorpresa aumenta.[277] La sorpresa mide lo improbable que era un suceso; la entropía mide lo improbable que se espera que sea (por término medio).[278] Así pues, la sorpresa, al igual que la entropía modelo, es algo malo para los organismos vivos. No quiero que nos perdamos en tecnicismos.[279] Permítanme solo decir que, al nivel realmente más básico, nos encontramos en un estado de sorpresa si salimos del conjunto de estados en los que biológicamente se prevé que estemos (por ejemplo, a más o a menos de entre 36,5 ºC y 37,5 ºC o respirando bajo el agua), precisamente porque hay una baja probabilidad de que nos encontremos en ese estado.

Los sistemas autoorganizados deben minimizar el flujo de información, porque el aumento de la demanda de información

implica un aumento de la incertidumbre en el modelo predictivo. La incertidumbre produce sorpresas, que son malas para nosotros en cuanto que sistemas biológicos porque pueden ser peligrosas. ¿Cómo minimizar las sorpresas reduciendo al mínimo el flujo de información? ¿No equivale eso a enterrar la cabeza en la arena? La respuesta es no. La energía libre de Friston es una medición cuantificable de la diferencia entre la forma en que un sistema modela el mundo y la forma en que el mundo se comporta realmente. Por lo tanto, debemos minimizar esta diferencia. El modelo del mundo de un sistema debe ajustarse todo lo posible al mundo real, lo que quiere decir que debe minimizar la diferencia entre los datos sensoriales que muestra y los datos sensoriales que había predicho su modelo. De esta forma, se maximiza la «información mutua» entre el mundo y el modelo, lo que minimiza la incertidumbre.

Una forma de hacerlo es mejorar el modelo del mundo que tiene el sistema. Los errores de predicción pueden ser retroalimentados al modelo generativo para que la próxima vez se puedan obtener mejores predicciones. Cuantos menos errores cometidos, menos errores para retroalimentar al modelo, lo que significa menor flujo de información. La «información mutua» es, pues, producto de la comunicación, de hacer y responder preguntas: ¿se comporta el mundo como yo predije, sí o no?

Ahora bien, como los sistemas biológicos como ustedes y yo estamos aislados del mundo por nuestras mantas de Markov, no podemos comparar nuestros modelos directamente con la forma en que el mundo es en realidad. Por lo tanto, debemos interiorizar en la cabeza todo el proceso de minimizar la sorpresa y convertirnos tanto en los «emisores» como en los «receptores» de la información que se deriva de nuestra formulación de preguntas.

Para ello, medimos entropías relativas, es decir, cuantificamos la diferencia entre los estados sensoriales predichos por una acción y los estados sensoriales que realmente se derivan de esa acción. Se obtiene así la cantidad llamada energía libre de Friston, que es siempre un valor positivo mayor que el valor de sorpresa real.

Las evidencias sensoriales (recibidas en forma de trenes de impulsos nerviosos generados dentro de nuestras cabezas, miles de millones de unos y ceros) son los únicos datos que podemos obtener. A partir de esos datos debemos inferir la estructura causal del mundo. Como los seres envueltos en mantas de Markov que somos, nos vemos obligados a confiar en cosas como las distribuciones de probabilidad, y no en verdades absolutas. Por eso es útil saber que la energía libre de Friston

es siempre mayor que el valor de la sorpresa; permite a nuestros cerebros aproximarse a verdades incognoscibles mediante cálculos estadísticos.

Como hemos visto en el experimento de la sopa de Friston, los modelos generativos surgen con los sistemas autoorganizados. Por esta razón, a veces hablamos de sistemas «autoevidenciables», porque modelan el mundo en relación con su propia viabilidad y luego buscan pruebas de sus modelos. Es como si en vez de decir «pienso, luego existo», dijeran «existo, luego mi modelo es viable».[280] El automodelo de cada sistema biológico viene determinado en parte por su especie, como ya he explicado. Nosotros, en cuanto que seres humanos, esperaremos encontrarnos en estados muy diferentes de los habituales para un tiburón. Es altamente improbable que nos encontremos respirando en el agua a cientos de metros bajo la superficie del mar, pero no es nada improbable que lo haga un tiburón. Así pues, un determinado conjunto de estados sensoriales es más o menos sorprendente —o más o menos improbable— según la especie de organismo que lo protagonice.

La prueba de un buen modelo del yo en el mundo es si permite al sistema del yo relacionarse con el mundo de forma que se mantenga dentro de unos límites viables. Cuanto mejores sean estas relaciones, menor será su energía libre. Cuanto menor sea su energía libre, más energía del sistema se consagrará a un trabajo efectivo de supervivencia.[281] El principio de la energía libre explica así, en términos matemáticos, cómo los sistemas vivos, como cada uno de nosotros, combaten la segunda ley de la termodinámica mediante un trabajo de mantenimiento de la homeostasis.

También explica otra forma en que los sistemas autoorganizados son autoevidenciables (casi diríamos introspectivos): están obligados a hacerse preguntas sobre sus propios estados. En concreto, se ven crónicamente obligados a preguntar: «¿Qué pasará con mi energía libre si hago tal cosa?». La respuesta a esta pregunta siempre determinará lo que el sistema haga a continuación durante un periodo de tiempo adecuado.[282] Este es el mecanismo causal que subyace en todos los comportamientos voluntarios.[283]

Sorprendentemente, todavía no hemos llegado al cabo de los notables poderes mentales que se aprecian en la sopa primigenia de Friston. Hemos visto cómo incluso los sistemas autoorganizados más básicos tienen un yo no consciente, así como subjetividad, agencia y un

propósito (es decir, sobrevivir). Perciben y, de ser necesario, actúan. A esta impresionante lista podemos añadir una habilidad más. Resulta que muestran una especie de racionalidad. Son, con una buena aproximación, bayesianos.

El reverendo Thomas Bayes fue un clérigo y teólogo inglés cuyo nombre sobrevive gracias a sus conocimientos sobre la probabilidad, que nunca se molestó en publicar. Bayes nos enseñó (en un artículo publicado póstumamente en 1763) que debemos utilizar las pruebas disponibles junto con los conocimientos previos para hacer y revisar nuestras mejores suposiciones sobre el mundo. En otras palabras, utilizando la terminología ya familiar de Friston, debemos tomar muestras sensoriales, compararlas con las predicciones derivadas de nuestros modelos generativos y actualizar nuestras creencias en consecuencia.

Esta es la expresión estándar:[284]

$$P(A \mid B) = [P(B \mid A)\, P(A)] \,/\, P(B)$$

Traducido en palabras, este teorema dice: «El coeficiente de probabilidades de dos hipótesis condicionales a un conjunto de datos es igual al coeficiente de sus probabilidades condicionales multiplicado por el grado en que la primera hipótesis supera a la segunda como predictor de los datos».

Más sencillo: dada una hipótesis seguida de algunas pruebas, es necesario revisar la probabilidad de la hipótesis considerando su verosimilitud en conjunción con su probabilidad anterior. La «verosimilitud» de la hipótesis es el grado de ajuste entre lo que predice y las pruebas realmente obtenidas, y la «probabilidad anterior» es el conocimiento previo que tenemos sobre la hipótesis (es decir, su probabilidad incluso antes de considerar las nuevas pruebas). El resultado es la probabilidad posterior de la hipótesis. Si partimos de dos creencias contrapuestas, nuestra mejor suposición es la que tiene la probabilidad posterior más alta (véase fig. 16, p. 239).

Por ejemplo, estamos en el aeropuerto de Ciudad del Cabo y vemos a alguien procedente de un vuelo de Johannesburgo que se parece a nuestra amiga Teresa. La hipótesis es que se trata de Teresa. Nuestra «probabilidad» es la probabilidad de estar viendo a alguien con ese aspecto, suponiendo que sea Teresa. Entonces recordamos que Teresa

vive en Londres, lo que reduce la «probabilidad anterior» de que sea ella. La conclusión (la «probabilidad posterior») es que estamos viendo a otra persona que simplemente se parece a Teresa.[285]

Lo más importante del teorema de Bayes a efectos de la neurociencia es que explica cómo funciona realmente en la vida real la inferencia perceptiva (un proceso inconsciente) y cómo funciona realmente la transmisión de señales en el procesamiento sensoriomotor real. Los circuitos cerebrales calculan literalmente distribuciones de probabilidades previas y luego envían mensajes predictivos a las neuronas sensoriales, en un esfuerzo sin fin por amortiguar las señales entrantes. La percepción implica literalmente comparaciones entre las distribuciones previsibles y las reales, lo que da lugar a cálculos de probabilidad posterior. Las inferencias resultantes son lo que es realmente esta percepción.[286] La percepción es un esfuerzo por autogenerar las señales sensoriales entrantes y así darlas por explicadas. Por eso muchos neurocientíficos hablan hoy en día del «cerebro bayesiano».

Recordemos lo que había dicho antes sobre la relación entre la «información» y el mundo material: aunque la información como tal no se pueda ver ni tocar, no cabe duda de que existe realmente. El comportamiento de los sistemas físicos viene determinado por los flujos de información. En consecuencia, la minimización de la energía libre de Friston minimiza al mismo tiempo la energía libre de Gibbs y de Helmholtz. Esto se debe a que minimizar el error de predicción minimiza el flujo de información, y reducir el flujo de información reduce el gasto metabólico del cerebro y del cuerpo en su conjunto.[287] Esto no solo se debe a que la actividad cerebral quema mucha energía (el 20 por ciento de nuestro suministro total). También se debe a que la minimización de la energía libre estadística en el cerebro regula los intercambios fisiológicos de energía entre el cuerpo y el mundo.[288] Vemos así que el cerebro predictivo es «perezoso» (a largo plazo): atento a cualquier oportunidad de conseguir más haciendo menos.[289]

Esta es una explicación minimalista de lo que hacen los cerebros. Sin embargo, para mantener vivo al organismo, el cerebro debe hacer algo más que conservar los recursos energéticos; también debe tener en cuenta muchas otras necesidades biológicas (además del balance energético) de las que hablo en el capítulo 5, que nos obligan, casi todas, a realizar un trabajo en el mundo exterior. Vemos así que la multiplicidad de necesidades que nos caracteriza a los organismos complejos tiene todo que ver con la conciencia.

Hemos aprendido que la supresión del error de predicción es el mecanismo esencial de la homeostasis. Por lo tanto, minimizar la energía libre se convierte en la tarea básica de todos los sistemas homeostáticos. La ecuación de la energía libre de Friston resulta ser una reformulación, en términos cuantificables, de la definición de «pulsión» de Freud: «medición de la demanda de trabajo que se hace a la mente a consecuencia de su conexión con el cuerpo»; una medición que Freud consideraba imposible de cuantificar. Ahora podemos cuantificarla. La pulsión fundamental del comportamiento volitivo de todas las formas de vida es que están obligadas a minimizar su propia energía libre. Este principio rige todo lo que hacen.

En palabras de Friston, el principio de la energía libre dicta que todas las magnitudes que pueden cambiar, es decir, que forman parte del sistema, cambiarán para minimizar la energía libre.[290] Llamemos a esto «ley de Friston»: todas las cantidades de un sistema autoorganizado que puedan cambiar cambiarán para minimizar la energía libre. Ya lo tenemos. Armados con este conocimiento, todo lo que llamamos vida mental se convierte en algo susceptible de ser matemáticamente procesado.

Estamos casi listos para saber qué es la conciencia, en términos formales y mecánicos; pero primero veamos una fábula.[291]

[228] Los cristales minimizan su energía libre de forma sencilla porque su estado estacionario de no equilibrio tiene un punto de atracción puntual. Es decir, simplemente se organizan en patrones compactos y allí se quedan, incluso cuando sufren ligeras perturbaciones. Las cosas se complican cuando el conjunto atractor tiene una estructura itinerante, con una dinámica como la que se ajusta al cerebro humano.

[229] Resumiendo, el índice h de un científico es el número de sus publicaciones que han sido citadas por sus pares más veces que la posición que ocupa esa misma publicación en la secuencia de publicaciones citadas. Así, si su cuadragésima publicación más citada ha sido citada cuarenta y dos veces, pero su cuadragésima primera publicación más citada solo ha sido citada treinta y nueve veces (es decir, menos de cuarenta y una), su índice h es cuarenta.

[230] A 20 de julio de 2020.

[231] Carhart-Harris y Friston, 2010.

[232] Freud, 1894, p. 60.

[233] Friston, 2005; la cursiva es mía. Lo mismo se puede decir de su artículo con Carhart-Harris, que comienza así (Carhart-Harris y Friston, 2010, p. 1265; la cursiva es mía): «Las descripciones de Freud de los procesos primarios y secundarios son coherentes con la actividad autoorganizada en los sistemas corticales jerárquicos y [...] sus descripciones del ego son coherentes con las funciones del modo por defecto y sus intercambios recíprocos con los sistemas cerebrales subordinados. Esta explicación neurobiológica se basa en una comprensión del cerebro como una máquina de Helmholtz o de inferencia jerárquica. Desde este punto de vista, las redes intrínsecas a gran escala ocupan niveles supraordenados de sistemas cerebrales jerárquicos que intentan optimizar su representación del sensorium. Esta optimización se ha formulado como la minimización de una energía libre; un proceso formalmente similar al tratamiento de la energía en las formulaciones freudianas».

[234] Friston, 2013.

[235] Solms y Friston, 2018.

[236] Parvizi y Damasio, 2003.

[237] Véanse Damasio, 2018; y mi revisión de este texto: Solms, 2018a.

[238] La idea de Einstein a la que me refiero es la naturaleza cuántica de la luz (Einstein, 1905). Stephen Hawking responde: «Dios no solo juega a los dados, sino que a veces nos confunde lanzándolos a un sitio en el que no se ven. Muchos científicos son como Einstein, en el sentido de que tienen un profundo apego emocional al determinismo. A diferencia de Einstein, han aceptado la reducción de nuestra capacidad de predicción que trajo consigo la teoría cuántica». Resulta que los fenómenos de la conciencia también son predecibles solo de forma probabilística.

[239] La escena tuvo lugar en la University of Southern California, en abril de 2018. En una llamada telefónica en enero de 2019, después de leer la versión publicada de nuestro artículo, seguía defendiendo «nuestra ciencia» sobre la base de que la conciencia era intrínsecamente biológica. Por fortuna ha cambiado de opinión desde

entonces; véase Man y Damasio, 2019.

[240] Es interesante saber que Hermann von Helmholtz (uno de los alumnos de Johannes Muller que fundó la Sociedad Física de Berlín) desempeñó un papel importante en la formulación de esta ley.

[241] Esto es cierto en la práctica, dentro del límite termodinámico, pero no (estrictamente hablando) en la teoría. Teóricamente sería más correcto decir: la segunda ley establece que es muy muy improbable que los procesos naturales sean reversibles. Esto se debe a que, cuando se resuelven las ecuaciones de cualquier sistema, no solo se tienen en cuenta las leyes dinámicas pertinentes, sino también las condiciones iniciales del sistema; y estas condiciones rompen la simetría de las ecuaciones fundamentales. Por lo tanto, si tuviéramos todas las piezas de una taza rota moviéndose unas hacia otras exactamente a la velocidad correcta, y las ondas sonoras moviéndose de vuelta hacia la taza exactamente de la manera correcta, y lo mismo sucediese con la energía que se hubiese disipado en el suelo cuando cayó, entonces todo podría volver a juntarse para recrear la taza inicial intacta. Sin embargo, como nunca nos encontramos en esa situación, nunca vemos decrecer la entropía en la realidad.

[242] Excepto durante periodos de tiempo muy muy breves.

[243] Cuando se añade calor a una sustancia, las moléculas y los átomos vibran más deprisa. A medida que los átomos vibran más deprisa, aumenta el espacio entre ellos. El movimiento y la separación de las partículas determinan el estado de la materia en la sustancia. Como resultado final del aumento del movimiento molecular, la sustancia se expande y ocupa más espacio.

[244] Técnicamente, la entropía se asocia a la energía desperdiciada, pero no son lo mismo. La entropía no tiene dimensiones; la energía sí.

[245] Técnicamente, se asocia al número de estados equivalentes correspondientes a una configuración macroscópica dada.

[246] El físico Alan Lightman (2018, pp. 67-68) lo expresa maravillosamente: «Si dividimos implacablemente el espacio en trozos cada vez más pequeños, como hizo Zenón, buscando el elemento más pequeño de la realidad, una vez que llegamos al mundo fantasmagórico de Planck, el espacio deja de tener sentido. Al menos, lo que entendemos por “espacio” ya no tiene sentido. En lugar de responder a la pregunta de cuál es la unidad más pequeña de materia, hemos invalidado las palabras utilizadas para formular la pregunta.

Tal vez así sea la realidad última, si es que existe. A medida que nos acercamos, perdemos las palabras».

Otro físico, Carlo Rovelli (2014, p. 167), ofrece un relato más prosaico: «El telón de fondo del espacio ha desaparecido, el tiempo ha desaparecido, las partículas clásicas han desaparecido, junto con los campos clásicos. Entonces, ¿de qué está hecho el mundo? La respuesta ahora es sencilla: [...] el mundo está hecho enteramente de campos cuánticos».

Incluso la incertidumbre cuántica forma parte del universo físico.

[247] Un «byte» consta de ocho bits. Por tanto, un «gigabyte» tiene 8.000 millones de bits. La velocidad de procesamiento de la información se expresa en «gigahercios»: un gigahercio equivale a mil millones de transferencias de bits por segundo. La realidad física de la información se refleja en el hecho de que estas unidades se pueden medir y también se pueden adquirir (a los proveedores de servicios de internet, por ejemplo).

[248] Véase el apéndice en la p. 367. Aquí hay que definir de qué incertidumbre se trata y para quién funciona. Lo mismo ocurre con «comunica» dos frases después: ¿quién se comunica con quién? Les invito a seguir leyendo.

[249] Gosseries et al., 2011.

[250] Si les parece que estamos confundiendo la entropía de la información con la entropía termodinámica, véase la cita de Tozzi, Zare y Benasich en la nota 61.

[251] Como se puede ver, el hecho de que tengamos múltiples necesidades, incluida la de BÚSQUEDA, es crucial.

[252] Consideremos, por ejemplo, el entrelazamiento cuántico entre dos partículas: una partícula «transporta información» acerca de la otra.

[253] Jaynes, 1957. Aumentar el tamaño del sistema descrito en nuestro ejemplo formal anterior (el gas en la cámara) aumenta su entropía termodinámica porque aumenta el número de posibles microestados del sistema que son coherentes con los valores mensurables de sus variables macroscópicas, lo que hace que cualquier descripción completa de su estado tenga más información. Para ser exactos, en el caso discreto que utiliza logaritmos de base dos, la entropía termodinámica reducida es igual al número mínimo

de preguntas sí/no que es necesario responder para especificar completamente el microestado, desde el momento en que conocemos el macroestado. Se puede encontrar una relación directa y físicamente real entre la entropía termodinámica y la entropía de la información asignando una unidad de medida a cada microestado que se produce por unidad de sustancia homogénea y calculando después la entropía termodinámica de estas unidades. Sobre la base de la teoría o de la observación, los microestados se producirán con diferentes probabilidades y esto determinará la entropía de la información. Así se demuestra que la entropía de Shannon es una medición estadística real de microestados que no tiene una unidad física fundamental distinta a las unidades de información.

[254] Shannon, 1948.

[255] Por ejemplo, el «principio de doble aspecto» de Chalmers, analizado en el capítulo 11.

[256] Wheeler fue alumno de Niels Bohr, quien formuló el principio de complementariedad. Este principio sostiene que los objetos tienen propiedades complementarias que no es posible observar y medir simultáneamente. Ejemplos de propiedades complementarias son la partícula y la onda, la posición y el momento, la energía y la duración, el giro en diferentes ejes, el valor de un campo y su cambio (en una posición determinada), y el entrelazamiento y la coherencia.

[257] La frase original es «It from Bit». He aquí la cita original (Wheeler, 1990, p. 5): «Todo sale de los bits. Dicho de otro modo, cada cosa (cada partícula, cada campo de fuerza, incluso el continuo espacio-tiempo) deriva su función, su significado, su propia existencia por completo —o indirectamente en algunos contextos— del sistema de respuestas a preguntas de sí o no, opciones binarias, bits. It from Bit simboliza la idea de que cada elemento del mundo físico tiene en el fondo (en un fondo muy profundo, en la mayoría de los casos) una fuente y una explicación inmateriales».

[258] Ibid. Las palabras citadas aquí vienen inmediatamente después de la cita anterior.

[259] Los lectores más atentos habrán observado que por fin me enfrento a la pregunta que me hice en mi infancia sobre el primer amanecer comentada en la introducción.

[260] Darwin, 1859. Véase Friston, 2013. Rovelli (2014, pp. 225-226) ofrece una lúcida explicación al respecto: «Un organismo vivo es un

sistema que se remodela continuamente, interactuando sin cesar con el mundo exterior. De tales organismos, solo siguen existiendo los que son más eficientes para hacerlo y, por tanto, los organismos vivos manifiestan propiedades que les han servido para sobrevivir. Por esta razón, son interpretables, y nosotros los interpretamos en términos de intencionalidad, de [objetivo y] propósito. Los aspectos finalistas del mundo biológico (este es el descubrimiento trascendental de Darwin) son, pues, el resultado de la selección de formas complejas eficaces para persistir, pero la forma eficaz de seguir existiendo en un entorno cambiante es gestionar mejor las correlaciones con el mundo exterior, es decir, la información; recoger, almacenar, transmitir y elaborar información. Por eso existe el ADN, junto con los sistemas inmunitarios, los órganos sensoriales, los sistemas nerviosos, los cerebros complejos, las lenguas, los libros, la biblioteca de Alejandría, los ordenadores y Wikipedia: maximizan la eficacia de la gestión de la información, es decir, la gestión de las correlaciones que favorecen la supervivencia».

[261] Lo que importa aquí no es que los sistemas autoorganizados estén siempre vivos (no lo están), sino que los sistemas vivos siempre son sistemas autoorganizados. Y lo que es más importante: no todos los sistemas autoorganizados son conscientes. En cuanto a la física de los orígenes de la vida, véase England, 2013, para un tratamiento interesante en una línea similar a la de Friston. La teoría de la «adaptación impulsada por la disipación» de England sostiene que los grupos de átomos impulsados por fuentes de energía externas tienden a aprovechar esas fuentes, alineándose y reorganizándose, para absorber mejor la energía y disiparla en forma de calor (es decir, minimizan su propia entropía a expensas de su entorno). Además, demuestra que esta tendencia disipativa favorece la autorreplicación: «Una buena forma de aumentar la disipación es hacer más copias de ti mismo». En cuanto a la biología del origen de la vida, véase Lane, 2015.

[262] Ashby, 1947. Véase también Conant y Ashby, 1970.

[263] Friston, 2013, p. 6. Lo que mostraba el monitor del ordenador eran, por supuesto, representaciones de los subsistemas, que no deben confundirse con la propia dinámica estadística. Lo mismo ocurre con los qualia perceptivos asociados al funcionamiento cortical, como veremos más adelante.

[264] Andréi Markov (1856-1922) fue un brillante matemático ruso, al igual que su hermano y su hijo. Trabajó principalmente en procesos estocásticos y hoy es famoso por lo que se conoció como «cadenas de

Markov» y «procesos de Markov». Era el perfecto rebelde. El rechazo institucional que sufrió fue tal que en vida nunca recibió reconocimientos académicos. En 1912 protestó contra la excomunión de Lev Tolstói de la Iglesia ortodoxa rusa solicitando su propia excomunión. La Iglesia accedió a su petición.

[265] Con «viceversa» no quiero decir que los estados internos del sistema no sean visibles para los estados externos (no pertenecientes al sistema), sino que el punto de vista del sistema está oculto; solo está disponible para sí mismo. Cuando se trata del problema de las otras mentes, un sistema nunca puede conocer los estados internos de otro sistema, no solo porque todos los estados externos están ocultos para él, sino también porque los estados internos de otros sistemas solo son internos respecto a esos sistemas.

[266] Friston, 2013, p. 8.

[267] El término autopoiesis fue introducido por Maturana y Varela (1972) para definir la química de mantenimiento autónomo de las células vivas.

[268] Teniendo en cuenta lo que he dicho antes sobre la capacidad de representación del sistema, los lectores filosóficos observarán que también aporta «intencionalidad» en el sentido de Brentano, 1874.

[269] Friston, 2013, p. 2.

[270] O simplemente está oculta y no está disponible para un trabajo eficaz.

[271] Es lo que Helmholtz llamaba la energía que no es libre (TS en la ecuación). La distinción entre energía «libre» y «ligada» seguramente no pasará desapercibida para los estudiosos de Freud. Véase el capítulo 10, nota 16, sobre la noción de Freud de «proceso secundario».

[272] $A = U + pV - TS$; donde p es presión y V es volumen. Esta ecuación cuantifica la energía libre de los sistemas cuyo trabajo está asociado a la expansión o compresión del sistema a temperatura y presión constantes.

[273] La ecuación tiene una forma larga y otra corta. Aquí está la forma larga: $F(s, \mu) = = E_q[- \log p(s, \psi \mid m)] - H[q(\psi \mid \mu)]$, donde F (energía libre «variacional», o energía libre de Friston para abreviar) es isomórfica de la energía libre de Gibbs y de la energía libre de Helmholtz. En cuanto a las demás magnitudes de la ecuación, s son los

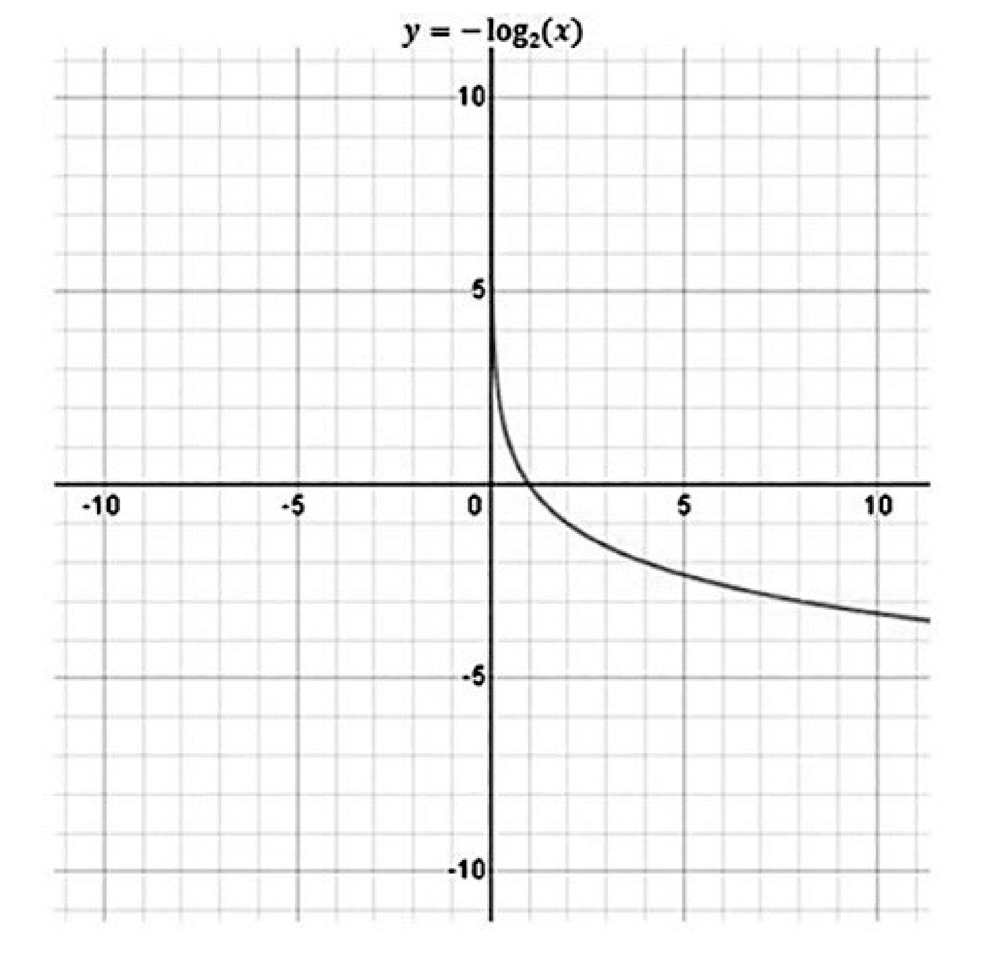
estados sensoriales (de la manta de Markov, comentada anteriormente), μ son sus estados internos, Eq es la energía media, p(s, ψ | m) es una densidad probable sobre los estados sensoriales y externos (ocultos) bajo un modelo generativo m, H es la entropía y q(ψ | μ) es una densidad variacional sobre los estados ocultos parametrizados por los estados internos. La relación entre esta ecuación utilizada en la ciencia de la información y la empleada en termodinámica no es evidente a primera vista. Sin embargo, cuando la ecuación larga se comprime así, se parece más a la termodinámica: F = Eq − H. Aquí, F es la energía libre de Friston, Eq es la energía media y H es la entropía.

[274] Es decir, la información media obtenida en muchas mediciones de microestados.

[275] Las demás expresiones citadas anteriormente (relativas a la sorpresa y la divergencia) nos dicen básicamente que, mientras que la energía libre de Helmholtz es una medida de la energía disponible para realizar un trabajo efectivo, la energía libre de Friston es una medida de la diferencia entre la forma en que un sistema modela el mundo y la forma en que el mundo se comporta realmente. (Dentro de un momento explicaré cómo se relaciona esto con el trabajo).

[276] El término de complejidad de la energía libre de Friston comparte el mismo punto fijo que la energía libre de Helmholtz (bajo el supuesto de que el sistema es termodinámicamente cerrado pero no aislado). Si las perturbaciones sensoriales se suspenden durante un periodo de tiempo convenientemente largo, la complejidad se minimiza porque se puede dejar de lado la precisión. En este punto, el sistema está en equilibrio y los estados internos minimizan la energía libre de Helmholtz por el principio de energía mínima (que es básicamente una reformulación de la segunda ley de la termodinámica).

[277] En otras palabras: la sorpresa es el logaritmo negativo de la probabilidad del resultado (s) según un modelo dado de los estados ocultos del mundo ([−log p(s, ψ | m)] en la ecuación anterior).



A medida que la probabilidad (eje x) se acerca a 0, aumenta la sorpresa (eje y); a medida que la probabilidad se acerca a 10, disminuye la sorpresa.

[278] Por lo tanto, minimizar la sorpresa sobre las cosas que han sucedido minimizará, por término medio, la entropía. (La sorpresa es un atributo de los datos o las observaciones, mientras que la energía libre es un atributo de las creencias. Por tanto, la parte de entropía de la energía libre no es la sorpresa media de las observaciones, es la entropía de las creencias sobre las causas latentes de las observaciones).

[279] La energía libre de Friston (energía media menos entropía) es equivalente a la sorpresa (que se expresa como $-\log p(s \mid m)$) más la «divergencia perceptiva» (que se expresa como DKL $[q(\psi \mid \mu) \mid\mid p(\psi \mid s, m)]$), que siempre es mayor o igual que la sorpresa por sí sola. La

«sorpresa» media es esencialmente entropía (de la información), como explico a continuación. La «divergencia perceptiva» mide la diferencia entre los sucesos hipotéticos y los reales según un modelo generativo. DKL es la divergencia perceptiva. Son las siglas de la divergencia de Kullback-Leibler, también conocida como entropía relativa, que cuantifica la divergencia entre dos densidades de probabilidad. Las dos que nos interesan se refieren a los estados ocultos: la densidad variacional codificada por los estados internos (por ejemplo, neuronales) y la densidad condicional real, dados los estados sensoriales. DKL es siempre mayor o igual a cero. Intuitivamente, esto se debe a que las funciones logarítmicas negativas (que se explican más adelante) siempre tienen tramas aproximadamente en forma de U, por lo que una línea que una dos puntos de la U nunca puede ser inferior a la parte inferior de la U (técnicamente, esto se denomina «función cóncava ascendente»). Así se garantiza que la energía libre imponga un límite superior a la sorpresa.

[280] Clark, 2017: «La autoevidencia […] se produce cuando una hipótesis explica mejor una prueba y, en virtud de ese éxito explicativo, aporta pruebas de su propia verdad o corrección. En estos casos, la hipótesis es la que mejor explica la aparición de las pruebas, pero el hecho de que se produzcan se utiliza para respaldar la hipótesis».

[281] Hay aquí otro giro técnico que revela la profunda relación entre autoorganización, teoría de la información e inferencia existencial. La sorpresa no es más que la probabilidad logarítmica negativa de algunos estados sensoriales para un modelo dado dentro de una manta de Markov. En estadística bayesiana, también se conoce como (el logaritmo de) la evidencia del modelo. Es lo que permite describir la minimización de la energía libre de Friston como autoevidencia; es decir, maximización de la evidencia para un modelo que genera los estados sensoriales de cualquier sistema que presente una manta de Markov.

[282] Por lo tanto, sería más correcto decir: «La respuesta a esta pregunta debe determinar el comportamiento medio del sistema».

[283] Atención: este mecanismo impone restricciones probabilísticas al «libre albedrío». Somos libres de entrar en las fauces de un león, por ejemplo, pero es poco probable que lo hagamos, y si lo hacemos, probablemente moriremos. Véase el teorema de Bayes, más adelante.

[284] Donde A y B son sucesos, $P(A \mid B)$ es la probabilidad de que ocurra A si B es verdadero (la probabilidad condicional), $P(B \mid A)$ es la

probabilidad de que ocurra B si A es verdadero (otra probabilidad condicional) y P(A) y P(B) son las probabilidades de que se den A y B independientemente una de otra (la probabilidad marginal). Mi traducción verbal del teorema sigue a Joyce, 2008, que escribió P(A) | P(B) como una relación separada.

[285] El teorema de Bayes puede reformularse en términos de energía libre, que se descompone en «precisión» y «complejidad». La evidencia del modelo es la diferencia entre precisión y complejidad, ya que los modelos con una energía libre mínima proporcionan explicaciones precisas de los datos a costa de la complejidad, lo que a su vez significa que reducir la complejidad del modelo mejora la posibilidad de generalizar el modelo pero a costa de la precisión. (En términos bayesianos, la «verosimilitud» debe evaluarse en relación con la «probabilidad» para evitar el sobreajuste). Knill y Pouget (2004, p. 713) llegan al centro de la cuestión: «La verdadera prueba de la hipótesis de la codificación bayesiana consiste en saber si los cálculos neuronales que dan lugar a juicios perceptivos o comportamientos motores tienen en cuenta la incertidumbre disponible en cada fase del procesamiento».

[286] Tenemos una exposición detallada de la microanatomía funcional de este proceso en Friston, 2005; Adams, Shipp y Friston, 2013; y Parr y Friston, 2018.

[287] Así, por ejemplo, los errores de predicción parecen comunicarse mediante frecuencias de rango gamma (altas), mientras que las predicciones parecen transmitirse mediante frecuencias de rango beta (bajas). Véanse Bastos et al., 2012, 2015.

[288] Véase Tozzi, Zare y Benasich, 2016: «Minimizar la energía libre variacional implica necesariamente una codificación metabólicamente eficiente que sea coherente con los principios de mínima redundancia y máxima transferencia de información (Picard y Friston, 2014). Maximizar la información mutua y minimizar los costes metabólicos son dos caras de la misma moneda; si descomponemos la energía libre variacional en precisión y complejidad, podemos derivar el principio de máxima información mutua como un caso especial de maximización de la precisión, mientras que minimizar la complejidad se traduce en minimizar los costes metabólicos (Friston et al., 2015). Así, la forma básica del principio de la energía libre de Friston refuerza la idea de que los niveles energéticos de la actividad cerebral espontánea, que son más bajos en comparación con la actividad evocada, permiten al sistema nervioso central obtener dos logros aparentemente contradictorios: minimizar al máximo los costes

metabólicos y, en la mayor medida posible, maximizar la información mutua».

[289] Véase Clark, 2015, p. 268. Sin embargo, esto debe sopesarse con lo que dice sobre la actividad por defecto del sistema de BÚSQUEDA (p. 263): «Estas criaturas están diseñadas para buscar pareja, evitar el hambre y la sed, y participar (incluso cuando no tienen hambre ni sed) en exploraciones esporádicas del entorno que les ayuden a prepararse para cambios ambientales inesperados, escasez de recursos, nuevos competidores, *etc.* Para cada momento, el error de predicción solo se minimiza con el telón de fondo de este complejo conjunto de "expectativas" que definen a la criatura».

[290] Friston y Stephan, 2007, p. 427.

[291] Basé esta fábula en la metáfora de Clark, 2015, de una presa con fugas de agua.

08

Una jerarquía predictiva

Una ingeniera de estructuras llamada Eva Periacueducto ha sido contratada para impedir y reparar las fugas en una presa municipal. No sabe que, a nivel más profundo, la verdadera finalidad de su trabajo es garantizar agua y electricidad a un pueblo de la zona e impedir que se lo lleve por delante una inundación. Pero no le hace falta saberlo; su única función es minimizar las fugas en la presa. (Tal vez recuerda de su época universitaria que lo que está minimizando es la entropía de la presa. Aunque tampoco necesita recordarlo; su trabajo es de carácter muy práctico).

La empresa que la contrata le facilita el equipamiento necesario junto con una pequeña brigada de trabajadores. También hereda un dosier de instrucciones redactadas por sus predecesores en las que se señala dónde están los puntos más débiles de la presa, diciéndole qué hay que hacer y cuándo. Eva y su equipo mantienen y reparan la presa con diligencia y actitud proactiva, centrándose en los puntos débiles, al tiempo que taponan las fugas espontáneas que van apareciendo. Con los años, Eva aprende que algunas de las fugas imprevistas también siguen patrones regulares, lo que la lleva a actualizar las instrucciones que le han pasado; así se vuelve más experta en la predicción (y, por ende, la prevención) de las fugas. Esto conlleva un ahorro de costes.

La emprendedora Eva se da cuenta de que los patrones de fuga a largo plazo que ha registrado mantienen una correlación con las condiciones meteorológicas. Sin querer, sus registros han modelizado el clima local (esto es, sus registros y la meteorología llevan «información mutua»). Sin pretenderlo, ha generado un modelo de un aspecto del mundo externo a la presa. En la meteorología hay patrones que se corresponden con los patrones de las fugas.

Ante este descubrimiento, Eva contrata a más personal para crear un departamento de meteorología, al que llama su departamento «sensor meteorológico». Se crea así un nivel nuevo en la jerarquía de su equipo, ubicado en otro lugar y justificado por la previsión de que disponer de mejores pronósticos meteorológicos le ahorrará a la larga costes de reparación.

El nuevo nivel hace su modelo predictivo más sensible a los contextos previsibles. Los miembros del equipo sensor meteorológico no

necesitan saber que su trabajo tiene algo que ver con impedir fugas; ellos solo se centran en la tarea de predecir los cambios de tiempo. Eva les facilita una tabla de condiciones previsibles derivada de las instrucciones heredadas con las que empezó y actualizada por ella. Obsérvese que esas instrucciones no son sobre patrones de fuga previsibles, sino sobre condiciones meteorológicas previsibles.

Dado que Eva no quiere perder demasiado tiempo comprobando mensajes, le pide al nuevo departamento que solo le envíe retroalimentación si se producen desviaciones con respecto a esas condiciones previsibles. A estos informes los llama «informes de error» y los utiliza para seguir actualizando su tabla de condiciones meteorológicas previsibles, que reenvía de vuelta a la estación, sabiendo que así reducirá la carga de trabajo de ellos y, en última instancia, la suya propia.

Todo esto permite al departamento de meteorología centrarse con eficacia en la tarea que les ha encomendado Eva. Para ello, instalan una serie de instrumentos de medición meteorológica en varias ubicaciones, algunas muy lejos de la presa. El equipo calibra los barómetros, termómetros y pluviómetros, entre otros, para que solo envíen señales a la estación meteorológica cuando los parámetros que recogen (presión atmosférica, temperatura, humedad, etc.) se desvíen de los límites previsibles. Estos límites se establecen en función de las condiciones meteorológicas predichas, lo cual vuelve a suponer un ahorro de costes porque los trabajadores contratados para leer los distintos medidores —con lo que se crea un tercer nivel en la jerarquía— solo tienen que verificar los instrumentos que transmiten señales de «error» a la estación meteorológica. Llevar registros minuciosos de esas señales permite a la estación ajustar con periodicidad los límites previsibles para cada instrumento, automatizando aún más sus procedimientos (esto es, reduciendo la frecuencia de las señales que los obligan a mandar a alguien a leer y ajustar los instrumentos).

Algunos de los algoritmos resultantes se vuelven bastante sutiles, a medida que el equipo va viendo que las fluctuaciones en los parámetros que miden no siempre son fijas y regulares; varían en función del contexto. Por ejemplo: «Si baja la presión atmosférica, aumenta la precipitación, pero solo en invierno». Con todo, los empleados que leen y ajustan los instrumentos no saben nada sobre la función más amplia del departamento de meteorología. Su único trabajo es leer mediciones y hacer los ajustes con arreglo a las instrucciones actualizadas que reciben desde la estación meteorológica. Y aún menos relevante para su trabajo es que los informes de su departamento sobre desviaciones respecto a las

condiciones meteorológicas previsibles se envíen a Eva para que pueda predecir con mayor fiabilidad los patrones de fugas y su trabajo de mantenimiento de la presa sea más eficiente.

De paso, los habitantes de los pueblos próximos utilizan esas previsiones del tiempo para sus propios fines, que no tienen nada que ver con la presa. Esto les da una idea equivocada de la verdadera finalidad del departamento, que creen que se ha creado para ayudarlos a programar actividades sociales al aire libre.

Con el paso del tiempo, Eva Periacueducto se da cuenta de que el patrón de fugas de la presa guarda una correlación no solo con las condiciones meteorológicas, sino también con episodios sísmicos. En consecuencia, crea un segundo equipo especializado, al que llama el departamento «sensor de terremotos». Este departamento de sismología se centra únicamente en modelar y predecir cambios tectónicos y afines, lo que lleva al segundo departamento a la instalación, calibrado, monitorización y ajuste constante de equipos «sensores» técnicos propios. También recopila registros complejos, lo que permite al nuevo equipo —como ya le pasó al departamento de meteorología y a la propia Eva— automatizar aspectos de su trabajo y centrarse solo en fluctuaciones impredecibles a corto plazo. (Lógicamente, los habitantes de los pueblos también aprovechan estas predicciones, aunque no fuera en ningún momento la intención original).

Lo que va emergiendo de forma gradual es una compleja jerarquía predictiva con múltiples departamentos, cada uno de los cuales tiene subniveles propios que muestrean distintos parámetros del mundo externo a la presa. Cada nivel de la jerarquía sigue solo las instrucciones predictivas actualizadas que recibe del nivel de encima, y solo informa de desviaciones respecto a los estados previsibles de los parámetros que se monitorizan en ese nivel. En lo que concierne a Eva, los informes combinados que recibe de sus departamentos sensores se contextualizan entre sí y ella tiene que decidir, de vez en cuando, qué informe conviene priorizar. Después de todo, sus recursos son limitados; no puede abarcar todos los sucesos posibles.

Eva continúa con el mantenimiento de la presa siguiendo su programa a largo plazo, del que solo se desvía cuando no encaja con las predicciones combinadas que recibe de sus departamentos sensoriales. Estos, a su vez, solo envían informes de retroalimentación a Eva cuando las muestras de datos que recogen se desvían respecto a sus predicciones formuladas hace tiempo. Y lo mismo va ocurriendo hasta llegar a las personas que leen y ajustan los medidores.

Por cierto, todo este intercambio iterativo de mensajes y esta actualización de programas entre los niveles de la organización creada por Eva sigue la regla de Bayes: utiliza las pruebas actuales (muestras sensoriales) junto con los conocimientos previos (hipótesis previa) para hacer y revisar sus mejores suposiciones (hipótesis posterior) sobre el mundo.

Conforme pasa el tiempo, el trabajo de Eva se vuelve repetitivo y aburrido, y ya piensa impaciente en la jubilación. Pero se le ocurre algo: «Antes de irme, me gustaría construir una presa completamente nueva y mejor». Y entonces llama al ayuntamiento y pregunta: «¿Por casualidad tenemos un departamento "reproductivo"?».

Muchas funciones cuasi mentales surgen con los sistemas autoorganizados más básicos, pero, para explicar el funcionamiento de los cerebros reales, hay que ver cómo pueden encajar estos sistemas en una estructura global y mutuamente beneficiosa. El trabajo de Friston nos ha permitido ver que el sistema nervioso implementa una jerarquía predictiva iterativa que funciona de manera muy similar a la que Eva Periacueducto fue creando a lo largo del tiempo. Las numerosas y complejas funciones del cerebro pueden, en última instancia, reducirse a unos pocos mecanismos sencillos como este. En palabras de Jakob Hohwy:

El cerebro intenta, de forma algo desesperada pero experta, contener los efectos a corto y largo plazo de las causas ambientales sobre el organismo a fin de conservar su integridad, lo que hace emerger implícitamente una representación del mundo rica y estratificada. Es una imagen bella y humilde de la mente y de nuestro lugar en la naturaleza.[292]

Cabe señalar que, en ciencia cognitiva, la palabra implícito significa «inconsciente».

En el centro del modelo que el cerebro tiene de sí mismo en el mundo, se generan predicciones propias de cada especie sobre sus límites viables (parte izquierda de fig. 15, p. 220).[293] Estas predicciones se encarnan en reflejos autónomos que adoptan la forma «si hago esto, mi temperatura será de unos 37 ºC». En el siguiente nivel (hacia la derecha), rodeando este núcleo, el cerebro genera comportamientos

instintivos (que adoptan la forma de las predicciones innatas que he descrito al hablar de las emociones básicas). En el siguiente nivel genera comportamientos involuntarios adquiridos (a partir de sus sistemas de memoria a largo plazo no declarativa). En el siguiente genera comportamientos voluntarios (a partir de sus sistemas de memoria a largo plazo declarativa). Y, al final, en el último nivel, genera las acciones más tentativas e inmediatas que «predicen el presente» (a partir de sus sistemas de memoria a corto plazo).

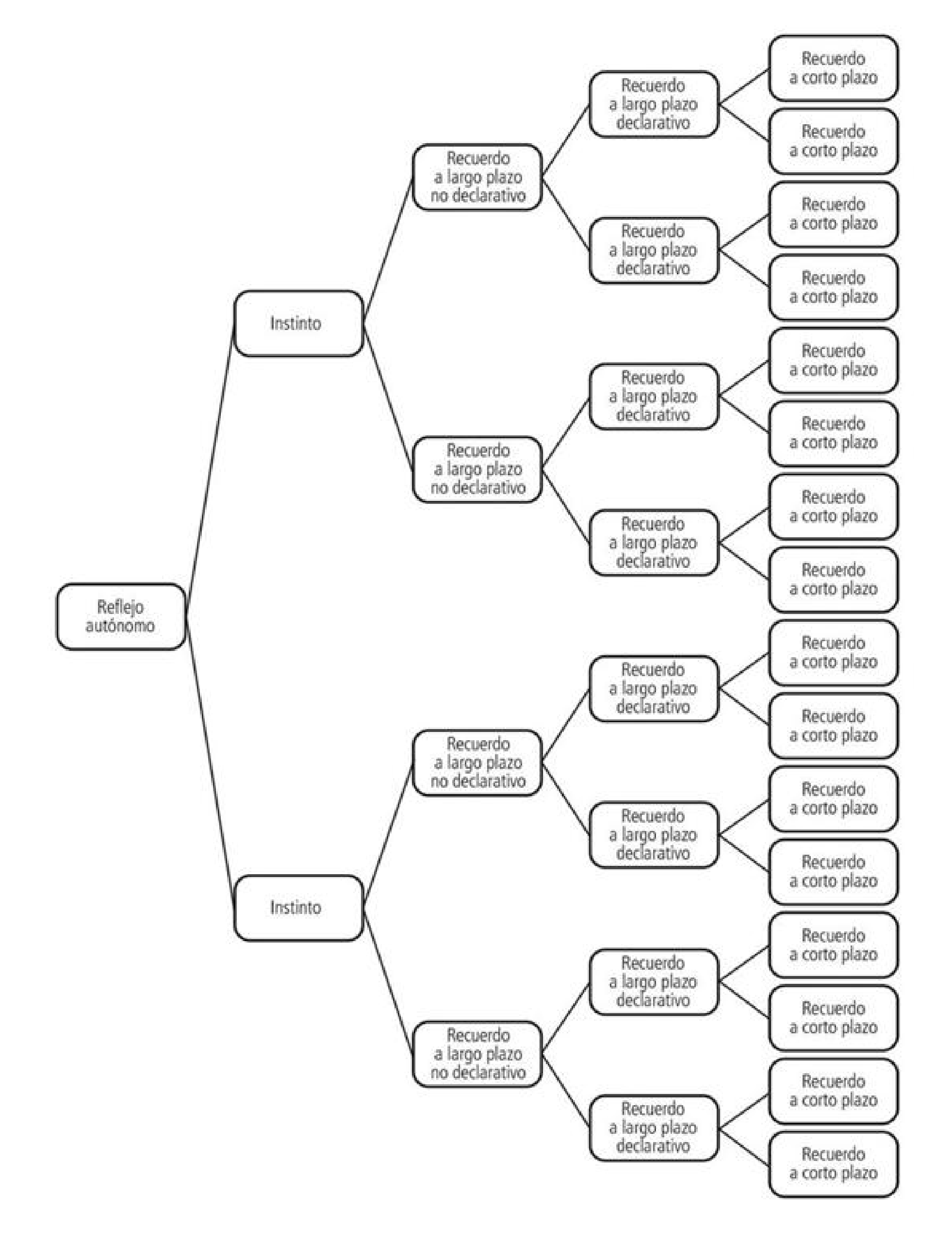


Figura 15

. Jerarquía predictiva simplificada, que va desde el núcleo autónomo tónico hasta la periferia sensoriomotora fásica. Las predicciones fluyen desde el núcleo a la periferia (de izquierda a derecha en este diagrama), en tanto que los errores de predicción fluyen en dirección contraria.

Es evidente que simplifico, porque en la jerarquía predictiva del cerebro hay muchos más niveles, no solo cinco, y se distribuyen en numerosos flujos de procesamiento paralelos. De todas formas, se pueden extraer varios principios generales.

El primer principio es que el cerebro conspira para anticiparse y «darnos explicados» los sucesos del mundo. Suprime las señales entrantes predecibles y poco informativas que, si no, tendría que procesar en vano. En resumen, cada nivel de la jerarquía recibe solo la información noticiable e inesperada que le transmite el nivel que está justo debajo. Estos informes de retroalimentación son errores de predicción.

El segundo principio general es que esta jerarquía se va desplegando siguiendo unas escalas temporales y espaciales cada vez menores. Las predicciones centrales son aplicables en todas las circunstancias, en tanto que las más periféricas son fugaces y focales. Una secuencia predictiva se desarrolla desde los núcleos de monitorización del cuerpo situados en el tronco encefálico y el diencéfalo,[294] pasa por los ganglios basales y el sistema límbico y, a través de la neocorteza, llega hasta los receptores sensoriales específicos de cada modalidad situados en los órganos terminales (por ejemplo, los conos y bastones de la retina), que tienen campos receptivos muy estrechos. En la periferia, la precisión y la complejidad a corto plazo prevalecen a costa de la generalizabilidad a largo plazo de la que gozan las predicciones más profundas.

El tercer principio guarda una estrecha relación con los anteriores: existe una jerarquía de plasticidad en virtud de la cual las predicciones centrales no pueden cambiar, pero las periféricas sí, y lo hacen; están sujetas a la actualización instantánea, y los niveles intermedios producen un grado intermedio de plasticidad. Esto significa que el «centro de control» del homeostato del cerebro (su automodelo) se actualiza constantemente, aunque presenta una resistencia cada vez mayor al cambio a medida que las cascadas de errores se acercan a su núcleo. La creciente plasticidad de los niveles más periféricos es una de las principales ventajas de un modelo predictivo jerárquico.

El cuarto principio, que hasta este momento no he explicado del todo, es que la percepción (por oposición al aprendizaje) invierte la dirección del procesamiento de la información. Al invertir las dependencias causales que han dado forma en un principio al modelo

predictivo, el cerebro produce nuestras inferencias perceptivas, que Merker describió como un «mundo tridimensional panorámico y completamente articulado compuesto de objetos sólidos con forma: el mundo de nuestra experiencia fenoménica familiar».[295] (Estas inferencias fluyen de izquierda a derecha en la figura 15). La «inversión del modelo predictivo» significa sencillamente pasar del aprendizaje a la predicción a partir de lo que hemos aprendido. Esto es lo que hizo Eva Periacueducto; infirió («percibió») el estado del mundo externo a la presa a partir de los datos meteorológicos y sismológicos que recibía.

La percepción procede de dentro afuera, siempre desde el punto de vista del sujeto. En realidad es apercepción, un proceso de inferencia, una cuestión de prueba de hipótesis bayesiana.[296] Hermann von Helmholtz, que fue quien primero entendió sus principios básicos, lo llamó «inferencia inconsciente» (atención otra vez al adjetivo). Lo que vemos es nuestra «mejor suposición» de lo que de verdad hay ahí fuera; es la respuesta que proponemos a las preguntas que le estamos formulando al mundo en ese momento.

El cerebro tiene que inferir las causas más probables de sus señales entrantes sin tener el menor acceso directo al mundo incognoscible que hay más allá de su manta. Lo único que tiene para seguir adelante es la forma en la que sus propios estados sensoriales (los trenes de impulsos nerviosos, fig. 11) fluyen y se modifican. Su tarea consiste en utilizar estas señales para crear un modelo probabilístico de las regularidades que existen en el mundo real (o más bien, entre él y el mundo) que luego utiliza para generar inferencias que guían sus acciones, las cuales tienen que asegurar su supervivencia en ese mundo. Asimismo, las acciones generan nuevas muestras sensoriales, que se utilizan para volver a actualizar el modelo, algo necesario porque los modelos son imperfectos. Esto conduce a más acciones, y así sucesivamente.

Así pues, deberíamos ver las acciones como experimentos que ponen a prueba las hipótesis derivadas del modelo generativo. Si un experimento no arroja los datos sensoriales previsibles, el sistema (1) debe cambiar su predicción para explicar mejor los datos o, si sigue confiando en la predicción original, (2) debe obtener mejores datos; es decir, debe realizar acciones que cambien su aporte sensorial.

Estas dos opciones —cambiar la predicción o el aporte— constituyen los mecanismos fundamentales de la percepción y de la acción, respectivamente.

Los tres párrafos anteriores me empujan a corregir un sesgo que ha caracterizado este capítulo y el que lo precede. Hasta ahora, como la mayoría de los científicos que utilizan las funciones corticales como su ejemplo paradigmático del funcionamiento del cerebro, me he centrado casi en exclusiva en la inferencia perceptiva bayesiana. Pero también existe la inferencia activa. De hecho, la inferencia activa es la forma principal (al menos en biología), ya que la razón de ser de la percepción es guiar la acción.

Como acabo de decir, el cerebro bayesiano tiene dos formas de responder al error de predicción. Ante una hipótesis en la que es aplicable la probabilidad «posterior» decreciente, ajusta mejor la hipótesis a los datos cambiando su predicción «previa» o su aporte. La diferencia entre ambas alternativas se reduce a la dirección estadística del ajuste: el error se reduce si la predicción se modifica para corresponder al aporte sensorial, y también se reduce si el aporte sensorial se modifica para corresponder a la predicción. En realidad, claro está, los organismos alternan todo el tiempo entre ambas opciones. (Pensemos en un ratón de campo que atraviesa corriendo la maleza, se para a mirar, se echa a correr otra vez, se para a mirar, y así sucesivamente). En algunos aspectos, la percepción y la acción son más similares de lo que parece.

El propio cuerpo es un mundo oculto «externo» a la manta de Markov de nuestro sistema nervioso central. La cascada de predicciones de la figura 15 podría haber incluido capas concéntricas equivalentes que culminasen en los receptores y los efectores viscerales terminales que hacen funcionar nuestros órganos internos. Así, el modelo del mundo que tiene el cerebro debe incluir un modelo de nuestro yo corporal y de su trayectoria en la misma medida que incluye todas las demás causas ocultas que nos interesan. (En el ejemplo ya comentado, no podríamos haber escapado de la habitación llena de humo si no tuviéramos un modelo implícito de la manera en que nuestro cuerpo se mueve y equilibra los gases en sangre).

Por otra parte, la acción no se produce porque el cerebro transmita un plan maestro a todos los músculos y órganos del cuerpo, sino porque los músculos se contraen y las glándulas secretan hasta que desaparecen los errores de predicción que transmiten a través de la jerarquía. Así pues, los órganos corporales musculoesqueléticos y viscerales de «acción» están a merced de las señales de error que generan las diferencias entre lo que el modelo predictivo espera que consigan y lo que en verdad consiguen. Suprimir el error de

predicción es lo que controla la acción, no menos que la percepción.[297]

Recordemos la ley de Friston: todas las cantidades de un sistema autoorganizado que pueden cambiar cambiarán para minimizar la energía libre. Los diversos homeostatos corporales regulados y orquestados por el metahomeostato del mesencéfalo son la clave del mecanismo que nos permite estar vivos, por la sencilla razón de que la regulación homeostática mantiene nuestro cuerpo dentro de sus límites viables. Estos límites no pueden cambiarse. Lo que significa —siempre siguiendo la ley de Friston— que tendrá que cambiar alguna otra cosa en el sistema. Esta es la explicación formal y mecanicista del vínculo imperativo que existe entre la pulsión y la acción, y es la razón de la necesidad de una jerarquía de predicciones previas, algunas de las cuales se pueden cambiar y otras no.[298]

Con todo, de nada sirve la acción si es ciega. Tiene que guiarse por la percepción, que es generada por un modelo del yo en el mundo. La regla de Bayes describe cómo se aplica el modelo predictivo que se encarga de esto, cómo se actualiza constantemente y por qué debe actualizarse. Esto pasa a ser la base formal del aprendizaje, por el que, con el tiempo, se van adquiriendo y matizando las predicciones a partir de las señales de error entrantes. Esta dinámica del sistema también confirma, por motivos igual de mecanicistas, que la acción se prioriza a la percepción: solo la acción puede incrementar las probabilidades de predicciones previas, algunas de las cuales, como he dicho, sencillamente no se pueden cambiar.

La regla de Bayes parte de una suposición de lo que llamamos «conocimiento de fondo». De lo contrario, la regla no puede funcionar.[299] Esto plantea una pregunta: ¿cómo se origina ese conocimiento de fondo al principio, cuando el sistema aún no ha reunido pruebas sobre el mundo? La respuesta es que nuestros principales «estados previsibles» están codificados por nuestra especie como puntos de estabilización homeostática innatos, cantidades determinadas por aquello que les funcionó bien a nuestros antepasados evolutivos. Somos beneficiarios de los éxitos biológicos de generaciones pasadas, que fijan las premisas más básicas de nuestra existencia.

Y no podemos dormirnos en los laureles. La conexión entre afecto y acción dicta que si al principio no logramos algo, debemos intentarlo una y otra vez. Las demandas de nuestras pulsiones biológicas más profundas son inexorables: solo pueden ser acalladas con su satisfacción o con la muerte. Si no se produce esto último, tenemos que complementar los reflejos e instintos con los que hemos nacido y

desarrollar otras formas de cubrir nuestras necesidades. No hay alternativa. En otras palabras, debemos aprender de la experiencia. Por suerte para nosotros, el cerebro humano está extraordinariamente bien equipado para ello.

Una implicación interesante de todo esto es que si el afecto funciona como lo hace mediante los mecanismos homeostáticos que he descrito —si realmente es «la medición de la demanda de trabajo que se hace a la mente a consecuencia de su conexión con el cuerpo»—, entonces tiene que ser el vehículo fundamental de minimización de la energía libre. El afecto, por consiguiente, es el medio principal de la volición, y la fuente de toda vida mental.

Antes he dicho que aprender de la experiencia produce un modelo jerárquico del mundo que si se invierte, genera predicciones sobre el mismo mundo. Sin embargo, el proceso no se acaba ahí; las predicciones hay que comprobarlas. Esto provoca errores de predicción que se utilizan para actualizar el modelo. En esto consiste el aprendizaje «a través de la experiencia». Los errores de predicción son las señales sensoriales que no ha predicho una hipótesis actual, es decir, las que no se han autogenerado. Esta es la parte destacada de los datos.

El error que comete la mayoría de los científicos cognitivos llegados a este punto es suponer que los datos entrantes son exclusivamente exteroceptivos. Olvidan que los errores de predicción (aportes sensoriales) que más nos importan llegan desde dentro. La desviación respecto a la temperatura corporal central previsible, por ejemplo, proporciona una retroalimentación «sensorial», tanto como los sucesos externos imprevistos. Lo mismo ocurre con la señal de error de homeostasis que da lugar a la alarma por asfixia. Estas señales generan afectos, no percepciones. Como dijo Freud, el prosencéfalo es un «ganglio simpático». Confundirse en este sentido es el eterno precio que pagan mis colegas por adoptar la falacia cortical.[300] La conciencia se genera endógenamente; toda ella. En su origen, la conciencia es afecto. Luego se extiende hacia el exterior, hacia la percepción, para evaluar las inferencias perceptivas, de la manera que describiré a continuación.

Por último, podemos abordar la pregunta de por qué y cómo se vuelven conscientes las funciones naturales de supervivencia descritas en este capítulo. Sabemos que la conciencia se fundamenta en el afecto, en las sensaciones y los sentimientos. Pero ¿cuáles son las leyes formales y mecanicistas que dan lugar a las sensaciones y los sentimientos y, por tanto, a la conciencia?

[292] Hohwy, 2013, p. 63.

[293] Digo «en el centro» en lugar de «en lo alto» de la jerarquía porque suena extraño (en términos anatómicos) ubicar los núcleos de monitorización corporal del hipotálamo y del tronco encefálico por encima de la neocorteza. Prefiero imaginar la jerarquía como un despliegue concéntrico, de dentro afuera, como las capas de una cebolla (véase Mesulam, 2000). Después de todo, como saben todos los embriólogos, el sistema nervioso es un tubo. Hay que recordar también que el tubo neural se forma a partir del ectodermo, mediante la invaginación de la placa neural, en donde el canal central ocupa el lugar del mundo exterior.

[294] Esta no es la visión estándar de la codificación predictiva, pero precisamente esa es la razón de este libro: demostrar hasta qué punto el corticocentrismo ha obstaculizado nuestra comprensión del cerebro. Cf. Pezzulo, 2014, p. 910: «Una dirección interesante para futuras investigaciones sería examinar la regulación homeostática en el marco de la inferencia activa».

[295] Clark (2015, p. 21) ofrece un breve resumen, utilizando la visión como su ejemplo modelo: «Un patrón de estimulación retiniana, que se dé en un contexto determinado, podría explicarse mejor utilizando un modelo generativo que […] combine representaciones fidedignas de agentes, objetos, motivos y movimientos que interactúen entre ellos y tengan múltiples capas intermedias que capten la manera en que los colores, las formas, las texturas y los bordes se combinan y evolucionan temporalmente. Cuando la combinación de esas causas ocultas (que abarcan numerosas escalas espaciales y temporales) se asienta en un todo coherente, el sistema ha autogenerado los datos sensoriales utilizando el conocimiento almacenado y percibe una escena significativa y estructurada. Cabe insistir de nuevo en que esta captura de la escena distal estructurada debe generarse utilizando solo la información disponible desde la perspectiva del animal. Tiene que ser una comprensión, estar completamente basada en la combinación de cualquier preestructuración (del cerebro y del cuerpo) que pueda estar presente gracias a la historia evolutiva del animal y a los juegos de estimulación energética que hayan registrado los receptores sensoriales. Un medio sistemático de alcanzar esa comprensión lo proporciona el intento continuo de autogenerar la señal sensorial mediante una arquitectura multinivel. En la práctica, esto significa que las conexiones descendentes y laterales dentro de un sistema multinivel llegan a codificar un modelo probabilístico de causas

interactuantes que operan a múltiples escalas de espacio y tiempo. Reconocemos objetos, estados y asuntos […] encontrando el conjunto más probable de factores interactuantes (causas distales), cuya combinación generaría (y, por tanto, predeciría y explicaría mejor) los datos sensoriales entrantes».

[296] Cf. Gregory, 1980.

[297] Este párrafo parafrasea a Hohwy (2013, pp. 81-82), pero corrige su sesgo exteroceptivo.

[298] No quiero decir con esto que los puntos de estabilización homeostática no cambien nunca, sino que sus grados de libertad son muy limitados; en consecuencia, casi todas las alostasis implican cambios de comportamiento, y no la actualización de predicciones anteriores innatas.

[299] Aquí paso por alto el tema del «Bayes empírico». Es posible aprender partiendo de cero, pero en términos biológicos resulta carísimo.

[300] Para ilustrar la magnitud del problema, véase Hohwy, 2013, p. 206: «A pesar de contar toda la historia de los errores de predicción, no parece haber contradicción alguna en concebir una criatura con toda la maquinaria para minimizar los errores de predicción que participe en la misma inferencia perceptiva que nosotros —en el grado de detalle neuronal y natural que decidamos especificar— y que sin embargo no sea fenoménicamente consciente. Esperaríamos que la criatura fuera consciente, claro, pero nada de toda la historia física implica que lo sea. Esto significa que queda abierta la cuestión de si la conciencia es algo que está por encima de lo físico o no».

¿Cómo podría una criatura con exactamente la misma maquinaria neuronal que ustedes y yo no ser consciente? Yo sostengo que el «zombi filosófico» que Hohwy imagina aquí carece de la maquinaria de los sentimientos. Para una visión alternativa del problema que plantea, véase el capítulo 12.

09

¿Por qué y cómo surge la conciencia?

La pregunta básica que todo organismo vivo debe hacerse siempre es: «¿Qué le ocurrirá a mi energía libre si hago eso?». Pero ¿hacer qué? Las acciones que se pueden emprender en un momento dado no son arbitrarias ni infinitas; las dictan las necesidades del momento. Hay, por supuesto, un vínculo íntimo entre necesidades y acciones; cada necesidad requiere una acción adecuada propia. Si tenemos hambre, debemos comer. Si estamos cansados, debemos descansar. Sin embargo, hay un cuello de botella ejecutivo: solo podemos hacer una o dos cosas a la vez. Eso significa que, para elegir la acción siguiente, hay que clasificar las necesidades del momento según la urgencia.

Son dos los factores que hacen que esa tarea de clasificación resulte más complicada de lo que podría parecer en un principio. En primer lugar, las necesidades de los organismos complejos como nosotros no tienen que ser satisfechas en un orden fijo. Comer y dormir…, ¿qué es más importante? Depende de toda clase de consideraciones. En segundo lugar, hay muchas necesidades de los organismos complejos que no siempre pueden ser satisfechas con la misma acción. Tomar sopa requiere destrezas distintas de las que se necesitan para comer una mazorca de maíz, por ejemplo (por no hablar de las destrezas y recursos necesarios para preparar la sopa y la mazorca). En ambos casos, lo «que toca hacer a continuación» depende crónicamente del contexto, y por ello hay que priorizar las necesidades en relación con las condiciones externas del momento.

En lo que respecta a qué hacer en estados como el hambre, la sed, el miedo y la rabia, nacemos con predicciones propias de la especie, unas predicciones innatas llamadas «reflejos» e «instintos» (herramientas de supervivencia heredadas por las que debemos sin duda estar agradecidos). Sin embargo, no son lo bastante flexibles para manejar la variedad y la complejidad de las situaciones a las que realmente nos enfrentamos; hay que complementarlas. Y ahí es donde entra el aprendizaje a través de la experiencia.

Ya hemos relacionado el aprendizaje a través de la experiencia con la ley del afecto. La valencia afectiva —nuestros sentimientos sobre lo que es biológicamente «bueno» o «malo» para nosotros— nos guía en situaciones imprevistas. Habíamos concluido que esa manera de

abrirnos paso entre los problemas imprevistos de la vida, sintiendo y mediante el comportamiento voluntario, es la función biológica de la conciencia. Ella guía nuestras opciones cuando nos encontramos perdidos en la oscuridad, pero, por supuesto, para que pueda hacerlo debe conectar nuestros afectos internos (arraigados en nuestras necesidades) con representaciones del mundo exterior.

Ese hecho explica por qué la excitación está acompañada de sensaciones y también percepciones conscientes de las cosas. Al final del capítulo 6 he reconocido que el gran misterio de esa conjunción —el misterio relativo a la manera en que la experiencia subjetiva encaja en el entramado del universo físico— solo podría desvelarse si reducimos los fenómenos fisiológicos y psicológicos por igual a sus causas mecanicistas subyacentes. Estas causas debían ponerse al descubierto a una profundidad de abstracción que solo la física podía proporcionar. En los dos capítulos anteriores he emprendido una explicación formal de esos mecanismos unificadores.

Ahora es el momento de completar esa explicación. Si la autoorganización y la homeostasis no explican por sí solas por qué y cómo surge la conciencia, ¿qué lo explica? ¿Cómo se relaciona, desde un punto de vista formal y mecanicista, el proceso biológico de priorización de necesidades que acabo de resumir con la minimización de la energía libre? ¿Y cómo es que el resultado de dicho proceso se siente como algo ante —y para— algunos sistemas autoorganizados?

El punto de partida de mi respuesta es precisamente el hecho que acabo de subrayar, a saber, que las criaturas complejas como nosotros, los vertebrados, tienen múltiples necesidades. Es decir, tenemos múltiples subsistemas internos, cada uno de los cuales está regulado por sus propios mecanismos homeostáticos, y todos ellos aportan valores de error al cálculo total de energía libre. Nuestras necesidades biológicas son esos valores de error. Cuando las necesidades empiezan a sentirse como afectos, decimos que están «valenciadas» positiva o negativamente, lo cual significa que tienen valor subjetivo: nos parecen buenas o malas. Los conductistas intentaron objetivar el valor redefiniendo los sentimientos de placer y displacer como estímulos de recompensa y castigo, pero este punto ya lo hemos tratado. La valencia no reside en el estímulo; es inherentemente subjetiva y cualitativa. Lo que para una persona es emocionante puede aterrorizar a otra.

Así y todo, ¿es imposible cuantificar la valencia? Si examinamos la

figura 12, vemos que cuanto más se desvía la flecha hacia la derecha, mayor es el displacer. Así pues, en un momento dado, nuestro valor de hambre puede ser 3/10 (que es peor que 1/10) y el valor de sed podría ser 2/10 (que es mejor que 5/10), y así sucesivamente. Los científicos afectivos siempre toman este tipo de mediciones y piden a los participantes de sus estudios que califiquen sus placeres y displaceres en las llamadas escalas de Likert. Aunque estas escalas son subjetivas, el hecho de que en principio sean cuantificables deja abierta la cuestión: ¿por qué se deben calificar los afectos? Si los sistemas autoorganizados pueden registrar como cantidades las «respuestas suscitadas por el equipo» (es decir, sus propios estados), entonces ¿qué añade la cualidad? Esta pregunta remite a lo que los filósofos llaman qualia, la elusiva materia mental que supuestamente no encuentra cabida en nuestra concepción fisicalista del universo.

Para hallar la respuesta, partamos del hecho de que no hay ninguna forma sencilla de combinar y totalizar las necesidades. Nuestras múltiples necesidades no pueden reducirse a un único común denominador; hay que evaluarlas con escalas separadas, más o menos iguales, de tal modo que a cada una se le asigne lo que le corresponde. No podemos limitarnos a decir que «3/10 de hambre más 1/10 de sed es igual a 4/20 de necesidad total» y luego intentar minimizar la suma total, ya que cada necesidad tiene que ser satisfecha por derecho propio. El metabolismo de la energía no es igual a la hidratación, que no es igual a la termorregulación, y así sucesivamente; cada una de ellas es una necesidad esencial. En palabras del científico conductista Edmund Rolls: «Si la recompensa en comida tuviera que ser siempre más intensa que otras recompensas, los genes del animal no sobrevivirían porque nunca bebería agua».[301]

Tras considerar todos estos factores, tiene sentido que los sistemas biológicos autoorganizados distingan sus necesidades (sus valores de error) categóricamente. La distinción entre variables categóricas es cualitativa. Dado que, por las razones que acabo de explicar, el error tipo A de 8/10 no se puede equiparar al error tipo B del mismo valor, hay que tratarlos como variables categóricas. Esto permite que a la larga el sistema asigne a cada uno de ellos lo que le corresponde y que también los priorice en su contexto. Por eso mismo es lógico que los sistemas autoevidenciables complejos categoricen (como «código de color» o «sabor») sus múltiples homeostatos a fin de poder computarlos con independencia unos de otros y priorizar los resultados.

No solo las distintas necesidades aportan distintas cantidades a la energía libre total; las distintas cantidades también tienen

implicaciones diferentes para el animal en contextos distintos (por ejemplo, el hambre supera a la somnolencia en algunas situaciones, pero no en otras). Esto contribuye, y no poco, a la incertidumbre, el enemigo mortal de las máquinas de predicción. Incrementar la incertidumbre conduce a un peligroso estado de cosas para cualquier sistema autoorganizado, pues predice la desaparición del sistema. Más incertidumbre exige más complejidad computacional (lo que significa más flujo de información y, a su vez, más entropía). Así, la categorización pasa a ser una necesidad cuando el valor relativo de distintas cantidades cambia a lo largo del tiempo (si ahora, pero no siempre, 8/10 para A vale más que 8/10 para B).

Es concebible que una serie sumamente compleja de algoritmos modelo llegue a evolucionar lo bastante para calcular demandas relativas de supervivencia en todas las situaciones predecibles, para permitirnos automáticamente priorizar acciones sobre esa base. No obstante, esos modelos complejos son muy costosos, y en todos los sentidos de la palabra. Son difíciles de manejar, lo que equivale a demora, que puede marcar la diferencia entre la vida y la muerte. Además, requieren muchísima potencia de procesamiento, lo que obliga a buscar más recursos energéticos. En estadística, al incremento exponencial de los recursos computacionales necesarios a causa de un incremento lineal de la complejidad del modelo se lo llama «explosión combinatoria».

Por otra parte, un modelo complejo que prediga con exactitud lo que ocurre en una situación específica difícilmente predecirá con la misma precisión lo que ocurrirá en otras situaciones. En términos estadísticos decimos que los modelos excesivamente complejos «sobreajustan» una muestra de datos. El modelo de predicción de fugas de Eva Periacueducto no estaba basado en hechos computados hora a hora y día a día durante unas cuantas semanas anteriores, sino que se basaba en promedios a largo plazo tomados de muchas muestras de datos recabados a lo largo de varios años, lo cual simplificaba su modelo y, por tanto, lo hacía más generalizable. Apoyándonos en el principio de la navaja de Ockham (ley de la parsimonia),[302] queremos modelos predictivos simples. La simplificación es esencial si nuestros modelos van a aplicarse en una amplia gama de situaciones. Deben ser utilizables no solo aquí y ahora, sino también en muchos otros contextos.

Insisto en que los modelos predictivos deben ser simples. Sin embargo, como dijo Einstein: «Todo debería hacerse tan simple como sea posible, pero no más simple».[303] ¿Cómo se alcanza el equilibrio justo? La compartimentación es el método estadístico clásico

empleado para conseguir el equilibrio óptimo entre complejidad y exactitud, y adopta muchas formas. Por ejemplo, una parte del sistema visual computa lo que miramos mientras otra parte computa dónde está, lo cual nos permite suponer la identidad constante de algo que se mueve a nuestro alrededor cambiando de forma, de tamaño y de orientación. Sin embargo, lo más importante es que la capacidad de compartimentar permite al sistema clasificar categóricamente, a lo largo del tiempo, sus necesidades y las predicciones concomitantes (es decir, las fuentes salientes de la energía libre previsible) y centrar sus esfuerzos informáticos en el compartimento priorizado.

Esa es la base estadístico-mecánica de un hecho observado, a saber, que cada afecto posee no solo una valencia hedónica continua (un grado de placer y displacer que es algo común a todos los afectos), sino también una cualidad categórica (así, por ejemplo, la sed se siente como algo distinto de la ansiedad por separación, que a su vez se siente diferente del asco, y así sucesivamente). Estos son los rasgos esenciales de los qualia afectivos, la forma elemental de todos los qualia: poseen cantidad y cualidad. Para decirlo con más claridad: los afectos son siempre subjetivos, valenciados y cualitativos; visto el problema que deben controlar —han evolucionado para hacerlo—, tienen que serlo.

Es útil describir en términos de modos operativos la selección y la priorización de categorías afectivas. Comparemos la manera en que se comporta un avión cuando está en los modos de despegue, vuelo de crucero, vuelo en turbulencias o aterrizaje. Las mismas variables entran en juego en estas distintas situaciones, pero tienen que ponderarse de manera distinta cada vez. Por ejemplo, la altitud exacta es mucho más importante durante el aterrizaje que durante el vuelo de crucero. Con la presa de Eva Periacueducto ocurre igual: los programas de funcionamiento variaban en los modos de invierno y de verano generados por Eva y en los modos terremoto y no terremoto. En condiciones de terremoto, Eva se vería obligada a cancelar el programa estacional normal y automatizado para aplicar un programa de «emergencia sísmica».

En los términos fisiológicos que empleo en el capítulo 6, los distintos modos operativos son funciones de estado del cerebro. Como explico en el citado capítulo, el triángulo decisorio del mesencéfalo selecciona estados cerebrales afectivos (como el modo «alarma por asfixia»). Lo hace cuando la SGP, la sustancia gris periacueductal, responde a esta pregunta: «¿Cuál de estas señales de error (léase «necesidades») convergentes brinda la mejor oportunidad para minimizar mi energía libre?». En otras palabras, ¿qué necesidad es la más prominente

(saliente) en este preciso momento? La respuesta la ofrecen no solo las magnitudes relativas de las señales de error opuestas, sino también las diferencias entre las categorías (modos o estados), cuya saliencia debe evaluarse en el contexto. Como ya he explicado, la información contextual la proporcionan los tubérculos cuadrigéminos superiores.

Veamos un ejemplo de algo que he visto hoy. Cuando he salido a correr a las siete, estaba oscuro, y a mi regreso al cabo de una hora, ya había amanecido. (Es invierno y estoy viviendo en el campo, en Sussex, donde escribo este libro). Al salir he pasado junto a un prado adyacente a la granja en el que había un rebaño de ovejas. Cuando han advertido mi presencia, casi se han caído unas sobre las otras por querer huir. A la vuelta, al pasar junto al mismo prado, las mismas ovejas, tumbadas en el mismo lugar, apenas me han mirado. El susto que se han llevado en la oscuridad ha pasado a ser indiferencia a la luz del día. En resumen, el contexto ha alterado la importancia del suceso «un humano se me acerca corriendo». De noche, el suceso es priorizado y pone a las ovejas en modo MIEDO; de día no, manteniéndose en su modo por defecto: BÚSQUEDA.

Este tipo de cosas determinan lo que un sistema hará a continuación. Dicho de otro modo, determinan qué estados activos serán seleccionados en el modelo generativo para resolver la categoría de incertidumbre priorizada. Es como si el sistema dijera: en las condiciones actuales, esta es la categoría de procesamiento de errores de predicción en el que no se puede sacrificar la complejidad informática. En consecuencia, el sistema (en este caso, las ovejas) se pone en modo operativo MIEDO y ejecuta la mejor estrategia que en las circunstancias presentes puede ofrecer su modelo generativo: huir. Después, tras tomar en cuenta todo lo que ha aprendido en el campo específico en que se encuentra (el contexto previsible), desea lo mejor pero se prepara para lo peor; con ello me refiero a que ejecuta con cuidado su plan, listo para adaptarse a los cambios que se produzcan.

Lo fundamental es que ponerse en modo MIEDO significa que la necesidad priorizada se ha convertido en un afecto. Dicho de otra manera, se ha vuelto consciente. ¿Por qué? Se vuelve consciente a fin de que las desviaciones de los resultados previsibles en la categoría de necesidad más saliente puedan sentirse a través de la jerarquía predictiva. Eso es el afecto, la «respuesta suscitada por el equipo» a la pregunta que el sistema se hizo sobre sí mismo: «¿cuál de estas señales de error convergentes brinda la mejor oportunidad de minimizar mi energía libre?».

En Sudáfrica, donde vivo la mayor parte del tiempo, la gente tiene

oportunidades de sobra para presenciar cómo funciona la selección de afectos en condiciones naturales; es decir, en el tipo de condiciones en las que este mecanismo evolucionó originalmente. No hablo aquí de lo que ocurre entre leones y gacelas en nuestras fabulosas reservas naturales (aunque ellos también brindan numerosas oportunidades). Me refiero a lo que ocurre en nuestra sociedad tan desigual y, por ende, con tantos conflictos. Muchos de mis compatriotas saben qué se siente cuando la necesidad más saliente es escapar de alguien que quiere matarnos. En ese momento, nuestro comportamiento voluntario es dominado por sentimientos de MIEDO, que evalúan el éxito o el fracaso aquí y ahora de nuestras acciones a medida que van desarrollándose. Otras necesidades (como la de orinar) quedan relegadas al automatismo. Dicho de otro modo, si nos hacemos pis encima no le daremos muchas vueltas.

Estamos tratando de reducir los fenómenos psicológicos y fisiológicos de la excitación afectiva a una serie de principios mecánicos que se pueden formalizar matemáticamente. En el capítulo 6 he explicado que una vez que el triángulo decisorio del mesencéfalo ha priorizado una necesidad, el modelo del prosencéfalo de su yo en el mundo genera un contexto previsible en el que esa necesidad será satisfecha. Asimismo, he dicho que ese mundo previsible tiene dos facetas: por un lado, representa el contenido real de nuestras predicciones; por el otro, debe codificar nuestro nivel de confianza en esas predicciones. La primera de dichas facetas la proporcionan las redes de recuerdos a largo plazo del prosencéfalo, que filtran el presente a través de las lentes del pasado. En el capítulo 7 he introducido los principios que rigen eso. La segunda faceta —el ajuste de los niveles de confianza— está modulada por la «excitación». Así pues, formalicemos las leyes que rigen dicho ajuste.

El primer mecanismo que he identificado hasta ahora es aquel por el cual se selecciona la categoría de acción más saliente (el modo de operación o el estado más eficaces). Así es como cualidades afectivas concretas llegan a regular por primera vez las acciones de los sistemas autoorganizados complejos, lo cual resulta en la generación de planos sensoriomotores, tras lo cual el sistema «desea lo mejor pero se prepara para lo peor». Y eso es parte de la historia en que me quiero centrar ahora. ¿Cuál es el mecanismo causal mediante el cual se regula ese desear lo mejor y prepararse para lo peor?

La primera parte de la respuesta dice que los niveles de confianza asociados a predicciones se aprenden a través de la experiencia, igual

que todo lo demás. Seguidamente, el nivel de confianza en nuestras predicciones se puede predecir, igual que las predicciones propiamente dichas. La codificación predictiva requiere que asignemos probabilidades a los estados sensoriales que esperamos se deriven de las acciones particulares y, luego, comparemos esas probabilidades con las distribuciones realmente observadas en las muestras sensoriales subsiguientes. Esa es la esencia del método bayesiano de «actualización de la hipótesis» que ya he descrito, el método con el que minimizamos nuestra energía libre.

Ahora bien, para determinar el ajuste entre un modelo y algunos datos no basta con comparar las medias de las distribuciones; es necesario también evaluar la variación respecto a las medias (véase fig. 16). En una muestra, una gran cantidad de variación hace que confiemos menos en el ajuste. Si una noticia dice: «El rey ha muerto, [...] el rey ha muerto, [...] el rey ha muerto», es más probable que la tome más en serio que si dice: «El rey ha muerto, [...] no, el rey no ha muerto, [...] al final, parece que sí, que el rey ha muerto». Los juicios de diferencia entre una distribución predicha y una muestra de datos son más fáciles de hacer cuando la distribución es limitada y precisa.

Nuestro objetivo ha de ser la precisión en nuestras interacciones con el mundo. En consecuencia, nuestros modelos deben tener un mecanismo para predecir la precisión. Así podremos «ponderar» la precisión esperada de las señales de error entrantes relativa a la precisión que asignamos a una predicción saliente. Esto (los grados relativos de confianza) dictará la influencia de las señales de error reales sobre nuestras predicciones. Si tenemos cada vez más confianza en una señal de error entrante, deberíamos empezar a confiar cada vez menos en nuestro plan de acción actual, en tanto que los cambios poco precisos e inesperados no deberían hacernos renunciar al curso que ya hemos determinado. (Los valores de confianza relativos pueden asignarse a las esperanzas activas frente a las perceptivas y las exteroceptivas frente a las interoceptivas también, y a todas las demás cantidades implicadas en la ley de Friston).[304] Se trata de una clase de inferencia bayesiana de segundo orden y conlleva inferencias sobre inferencias, esto es, niveles de confianza bien instruidos sobre las predicciones.

La finalidad de la modulación de la precisión es garantizar que las inferencias hechas por los modelos predictivos son impulsadas por señales de aprendizaje fiables (noticias dignas de confianza): si hay mucha confianza en una señal, debería permitirse revisar una hipótesis previa, y si hay poca, no debería permitirse. La confianza afecta a la intensidad de las señales de error que se propagan hacia

dentro a través de la jerarquía. Una señal en la que tenemos mucha confianza (esto es, una señal más precisa) es una señal «más alta». Por tanto, tendrá más oportunidades de ofrecer algún error residual al núcleo del sistema y más oportunidades de actualizar su modelo generativo. A la inversa, las señales menos precisas —señales en las que uno tiene menos confianza, también llamadas «ruido»— pueden acabar secuestradas en el epitelio sensorial y (lo que sería de esperar) siendo ignoradas para más seguridad.

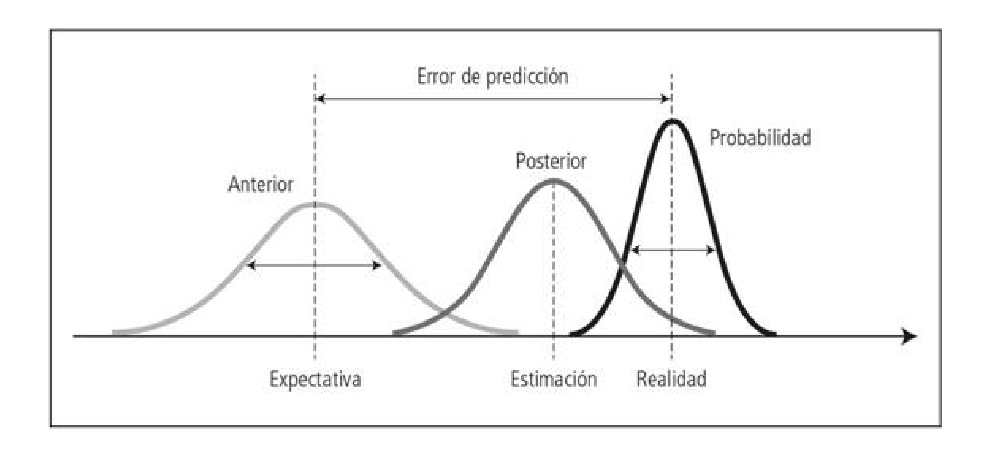


Figura 16

. Una predicción anterior (a la izquierda) se compara con una muestra de datos sensoriales (a la derecha), lo cual resulta en una predicción posterior (centro). Las «medias» de estas tres distribuciones se indican con las líneas (de puntos) verticales, y su «varianza», con las horizontales (flechas). La amplia varianza previsible en la distribución anterior (línea horizontal a la izquierda) indica un bajo grado de confianza en la predicción anterior, y la estrecha varianza real en la muestra de datos sensoriales (línea horizontal a la derecha) indica un alto grado de confianza en los datos. En este ejemplo, la precisión (varianza inversa) de los datos sensoriales es alta; a resultas de ello, la predicción posterior pasa claramente a la derecha. Si la precisión de los datos sensoriales fuera más baja, la predicción posterior se movería menos (en caso de que se moviera).

Esto significa que debemos minimizar las señales de error precisas. Una vez más, suena a paradoja hasta que nos damos cuenta de que solo significa que tenemos que evitar cometer errores flagrantes. La

única manera de conseguirlo es mejorar nuestros modelos generativos aumentando así la información mutua entre nuestros modelos del mundo y las muestras sensoriales que obtenemos de él. En otras palabras, tenemos que maximizar la precisión de nuestras predicciones y luego buscar datos precisos que las confirmen. Debemos maximizar nuestra confianza en las creencias que guían nuestras acciones. A esto se lo llama «optimización de la precisión».

Dicha optimización de la precisión se logra aprendiendo de la experiencia. Tenemos que aprender en qué fuentes de noticias podemos confiar (y cuándo) y ajustar nuestras predicciones en consecuencia. Por ejemplo, confiamos en nuestras señales visuales diurnas y en las auditivas por la noche. Asignamos mayor peso a lo que vemos de día y a lo que oímos de noche porque hemos aprendido a hacerlo así. Por consiguiente, esperamos que la información visual sea más precisa de día que de noche. (Esto se llama «precisión esperada»). Dado que no esperamos precisión visual por la noche, toleramos sin más toda clase de imágenes vagas y borrosas; pero experiencias visuales similares de día nos harían pensar que hay algo que va muy mal.

Del mismo modo, el personal de Eva Periacueducto aprendió que la presión barométrica a la baja predice un aumento de la precipitación en invierno, pero no en verano. Por lo tanto, los empleados de Eva podrían tratar una lectura a la baja como «ruido» con más seguridad en verano que en invierno. Esperarían que las lecturas barométricas fuesen menos precisas en verano. Si no fuera así, adaptarían sus niveles de precisión esperada, algo que a su vez afectaría a las predicciones posteriores. Por tomar un ejemplo más afectivo: aquellas ovejas de Sussex aprendieron a confiar en la predicción de que la gente que corría hacia ellas durante el día no podía hacerles daño, pero desconfiaban más de esa predicción por la noche. Por consiguiente, le asignaban menos precisión esperada durante la noche. Es posible que si volvieran a encontrarse conmigo más veces en la oscuridad y constataran que no sufren daño alguno, ajustaran ese valor de precisión y modificaran en consecuencia sus predicciones respecto de la gente que corre hacia ellas de noche. Las dependencias contextuales pueden aprenderse igual que todo lo demás.

A partir de lo que he dicho, está claro que esa clase de aprendizaje gira en torno a contextos variables. Sin modelar predictivamente la dinámica contextual, un sistema autoorganizado no puede sobrevivir mucho tiempo en entornos cambiantes. El modelo generativo está obligado a incorporar esas dinámicas; tiene que aprender a predecir grados de precisión. Y el ajuste de los valores de precisión, como todo

lo demás en el cerebro predictivo, debe seguir la ley de Friston.

La precisión es el modo en que el cerebro representa su grado de confianza en una fuente dada de evidencia sensorial o en las consecuencias predichas de una acción dada. Los valores de precisión cuantifican las expectativas relativas a la variabilidad. Son, por tanto, representaciones de incertidumbre. ¿Cuánto confío en esta señal de error en el contexto actual? ¿Cómo debería ponderarla en este preciso momento? ¿8/10 para A vale más o menos que 8/10 para B en las condiciones actuales?

Ya hemos visto el aspecto fisiológico de este punto: el triángulo decisorio del mesencéfalo prioriza una necesidad; luego, el modelo del prosencéfalo del yo en el mundo genera un contexto previsible en el que se satisfará la necesidad priorizada. Ese mundo previsible tiene dos facetas, a saber, el contenido real de las predicciones y el nivel de confianza que el sistema deposita en esas predicciones. Ahora que sabemos cómo se cuantifican los niveles de confianza, podemos introducirlos en nuestra explicación de la fisiología de la excitación.

Una vez que el triángulo decisorio ha seleccionado su necesidad saliente actual —la que determina el estado afectivo del sistema, que a su vez determina el contexto previsible que generan los sistemas de recuerdos a largo plazo del prosencéfalo—, empieza a trabajar el sistema reticular activador. Los sistemas de memoria asignan valores de precisión de referencia para el contexto previsible y los aplican a toda la jerarquía predictiva. A continuación, una nube de neuromoduladores atraviesa el prosencéfalo instando a algunos canales a que disparen rápido y desalentando a otros. Las tasas de disparo determinan cuánto peso se asignará a las predicciones actuales y a sus errores concomitantes, que regirán «cuán alto» se transmitirán los errores. En otras palabras, los valores de precisión determinan cuánto confía el sistema en los resultados que espera que se deriven a partir del curso de acción que se desarrolla ahora para todos los distintos niveles de la jerarquía. Y, una vez más, desea lo mejor pero se prepara para lo peor.

Son muchas las cosas sorprendentes que pueden predecirse una vez nos acostumbramos a ellas. No obstante, si desear lo mejor y prepararse para lo peor es lo único que podemos hacer, esto implica que hay algunas cosas que no se pueden predecir. Esa es la segunda parte de la historia, y requiere que el sistema ajuste sus niveles de confianza sobre la marcha (es decir, que module la excitación en el

contexto de los sucesos tal como se desarrollan).

Antes he dicho que es concebible que una serie sumamente compleja de algoritmos modelo llegue a evolucionar lo bastante —con independencia de lo poco manejables que lleguen a ser— para calcular demandas relativas de supervivencia en todas las situaciones predecibles, para priorizar su acción sobre esa base, a pesar de la «explosión combinatoria». Sin embargo, ¿cómo escoge el organismo entre A y B cuando la incertidumbre misma pasa a ser el determinante principal de la selección de acciones? Eso es lo que ocurre, por ejemplo, cuando se producen situaciones nuevas, algo nada excepcional en la naturaleza.

Lo que los fisiólogos llaman «modulación de la excitación», para los científicos computacionales es «ponderación de precisión»;[305] son una y la misma cosa. Como acabamos de ver, una señal precisa no es más que lo que en el capítulo 6 llamo una señal «alta» —una señal intensa—, y eso implica que la modulación de la confianza en una señal de error debe seguir las desviaciones de su intensidad previsible. Esas desviaciones deben minimizarse. Como ocurre con todas las señales de error de homeostasis, es «buena» (para nosotros, sistemas biológicos) cuando las cosas salen como se esperaba y «mala» cuando prevalece la incertidumbre.

A medida que se desarrolla la secuencia de acciones, los niveles de confianza de referencia se ajustan al alza y a la baja por medio del sistema reticular activador. (Pensemos en los empleados de Eva Periacueducto encargados de leer y ajustar los instrumentos). Es decir, que el contexto sensoriomotor en curso se «palpa» —y se ajustan las ponderaciones de confianza del sistema— sobre la base de fluctuaciones en curso de la incertidumbre previsible. Las alteraciones de la excitación rastrean la fiabilidad estimada de los errores de predicción muestreados. De este modo, los valores de precisión variables estiman la fiabilidad cambiante de las señales en curso que transportan la noticia. A su vez, dichos valores determinan todo lo que además hace el sistema, de acuerdo con la ley de Friston.

Todo lo dicho hasta aquí sugiere que la optimización de la precisión es la base estadístico-mecánica de la priorización de señales en general, es decir, el resultado crítico de todo lo que hemos visto en el triángulo decisorio del mesencéfalo y el sistema reticular activador. La optimización de la precisión es la manera en que las múltiples señales de error que convergen en la SGP se priorizan en primer lugar, llevando la necesidad más saliente a la conciencia afectiva y dando lugar a una serie de opciones que se desarrollan en un contexto

previsible guiadas por las precisiones esperadas, que ahora deben ser moduladas a partir de los sucesos sensoriales imprevistos.

Es posible que esto suene algo abstracto. Personalmente pienso que, al contrario, es muy fiel a la vida cotidiana. Gran parte de nuestra experiencia se reduce a breves avisos de sensaciones cuando detectamos cosas que no son exactamente como esperábamos que fuesen, tras los que hacemos una exploración cognitiva en busca de formas de cerrar esa brecha. De pronto nos acordamos de que debemos enviar un correo electrónico: solo cuando la mano no detecta la superficie del móvil tomamos conciencia de que estábamos intentando cogerlo (pero si no está justo ahí, a nuestro lado, ¿dónde lo hemos dejado? ¿En la cocina, donde estábamos hace cinco minutos?).

Veamos otro ejemplo, bastante más importante desde el punto de vista afectivo: un encuentro con un o una posible amante. Pensamos que podríamos seducirlo o seducirla esta noche, imaginamos la posible secuencia de los hechos y fijamos un plan de acción. Después deseamos lo mejor. No estamos seguros de cómo saldrán las cosas, pero, basándonos en nuestra experiencia previa con esa persona, consideramos que tenemos alrededor de un 7/10 de probabilidades.

A medida que va transcurriendo la velada, nuestro foco de atención (lo saliente) es bastante distinto del que sería si estuviéramos cenando con nuestro hermano. El tono emocional también es distinto. Cada pequeño signo que sugiere que la persona sentada al otro lado de la mesa reacciona positivamente a nuestros lances provoca ráfagas de emoción, y así aumenta la confianza en que nuestro plan funciona. De repente, la persona deseada bosteza y mira la hora. ¿Qué significa eso? ¿Cuánto peso habría que asignarle a ese detalle? Tenemos un mal presentimiento. ¿Hemos leído mal todos los signos anteriores? Estudiamos cada movimiento y cada gesto; la menor nueva indicación de que nuestros sentimientos no son correspondidos nos preparará para lo peor y pondremos en marcha el plan B: salvar nuestro orgullo fingiendo que somos igual de indiferentes. Pero de pronto las miradas se cruzan. ¿Significa eso lo que pensamos que significa? ¡Sí! Y en ese momento sentimos que nos tocan cariñosamente la mano. El corazón se acelera. Parece que, después de todo, podemos seguir fieles al plan A.

Una necesidad priorizada (en este caso, la LUJURIA) es en esos momentos la fuente de incertidumbre más saliente. Las inferencias sobre sus causas se vuelven conscientes como afecto, porque las fluctuaciones del nivel de confianza en cuanto a las posibles acciones necesarias para satisfacer esa necesidad deben modularlas los

sentimientos. Son ellos los que nos dicen si estamos haciendo las cosas bien o mal. El contexto en curso que da lugar a las fluctuaciones también debe, por la misma razón, volverse consciente. Ese es el motivo por el que he definido la conciencia exteroceptiva de acción y percepción como afecto contextualizado. Ahora tenemos una comprensión formal y mecanicista de lo que eso significa. Todo es solamente incertidumbre sentida.

Es de fundamental importancia señalar que la afirmación «el contexto en curso que da lugar a las fluctuaciones también debe [...] volverse consciente» explica por qué la experiencia tiene aspectos duales. No es solo una cuestión de «me siento así», sino más bien de «me siento así respecto de eso». El «respecto de eso» también tiene que ser sentido empleando una divisa común (la incertidumbre aplicada), porque el contexto es la fuente principal de la incertidumbre sobre la energía libre. La economía de la minimización de la energía libre reclama una divisa común.

Estos hechos revelan que la conciencia no es meramente una perspectiva subjetiva de la dinámica «real» de los sistemas autoorganizados, sino una función con poderes causales definidos propios. La sensación de una necesidad (como opuesta a la mera existencia de una necesidad) determina, y mucho, lo que el sujeto de esa necesidad hará luego. Los afectos impulsan literalmente lo que un animal hace de un momento a otro en condiciones de incertidumbre. La finalidad de las percepciones exteroceptivas consiste en que sean sentidas en relación con las acciones impulsadas afectivamente que contextualizan.

Esa es la función principal de elementos como la atención. El foco de atención funciona como la selección de afectos, pero se aplica al mundo exterior. Nuestra necesidad de reducir la incertidumbre gobierna nuestra mirada, por ejemplo, de tal modo que los movimientos oculares rápidos rastrean las regiones de una escena donde hay más probabilidades de encontrar información más precisa.[306] Dicho de una manera más sencilla, las señales relativamente intensas atraen la atención. Se les asigna una precisión más alta.

Así trabaja la saliencia. Los rasgos «salientes» del mundo son rasgos que, cuando se muestrean, minimizan la incertidumbre relativa a la hipótesis del sistema actualmente priorizada; son ellos los que, cuando las cosas van como se esperaba, maximizan nuestra confianza en la hipótesis, lo que lleva a los agentes activos a tomar muestras del mundo con vistas a (intentar) confirmar sus propias hipótesis.[307]

Dado que estas últimas, en los análisis finales, son hipótesis acerca de cómo satisfacer nuestras necesidades, esto significa que cada especie se ve impulsada a seleccionar su propio mundo perceptivo. La orientación perceptiva de cada especie la dictan las cosas que le importan. Por consiguiente, humanos, tiburones y murciélagos vivimos en distintos mundos (subjetivos). Los objetos y los sucesos solo se perciben cuando se ven, y cada especie tiene sus objetos y sucesos salientes. Solo se pueden ver los que se muestrean.[308] El biólogo chileno Francisco Varela lo dice con bonitas palabras: «la especie crea y especifica su propio ámbito de problemas».[309]

Esto implica que la precisión no se puede determinar pasivamente. No podemos limitarnos a esperar hasta ver qué señales son intensas sin expectativa alguna en ningún sentido; eso es algo que el modelo generativo debe inferir y luego asignar. La atención —que tiene todo que ver con la precisión— puede, en consecuencia, ser «atrapada» y «dirigida».[310] Por ejemplo, cuando vamos a dormir, reducimos activamente casi a cero la precisión en los errores sensoriales, pero un suceso lo bastante sorprendente aún podrá despertarnos. En otras situaciones podríamos amplificarlo activamente, como cuando nos concentramos en un texto denso porque sospechamos que contiene algo importante.

En el artículo de 2018 en que expusimos nuestra teoría de la conciencia, Friston y yo alteramos algunos de los símbolos convencionales empleados en las ecuaciones que calculan la energía libre (véanse las notas del capítulo 7) y los reemplazamos para reconocer que seguimos los pasos de Sigmund Freud, quien en 1895 intentó «estructurar una psicología que sea una ciencia natural: es decir, representar los procesos psíquicos como estados cuantitativamente determinados de partículas materiales especificables, haciendo así que esos procesos se vuelvan intuibles y libres de contradicción».[311]

Esas fueron las primeras líneas de su «Proyecto para una psicología científica». En este texto, Freud utilizó los símbolos φ, ψ, ω y M para referirse a cuatro sistemas hipotéticos de neuronas responsables de la percepción, la memoria, la conciencia y la acción, respectivamente, y empleó el símbolo Q para indicar los estímulos externos. Friston y yo seguimos este uso para indicar los vectores equivalentes dentro de un sistema autoevidenciable:[312]

Qη = estados externos, tal como los modelan los estados internos del sistema

φ = estados sensoriales

M = estados activos

ψ = predicciones

ω = precisiones

Asimismo empleamos:

e = errores de predicción (basados en φ y su predicción ψ)

F = energía libre de Friston (basada en e y precisión ω).[313]

Obsérvese que ninguna de estas cantidades mide directamente estados externos, puesto que a un sistema autoorganizado se le ocultan. En todo lo que expondremos en adelante, eso significa que tenemos una descripción independiente y autónoma de la dinámica mental en lo que respecta a los propios estados internos del sistema (Qη, ω) y la manta de Markov (φ, M). Equipados con estos términos, ahora podemos formalizar una dinámica del sistema autoevidenciable en relación con la optimización de la precisión.

Empezaré con dos ecuaciones que definen la energía libre de Friston en relación con las cantidades que he introducido. La primera ecuación dice que la «energía libre es (aproximadamente) el logaritmo negativo de la probabilidad de encontrar algunos estados sensoriales activamente creados».[314] La segunda dice que «la energía libre previsible disminuye en proporción (aproximada) a la precisión del logaritmo negativo».[315] Recordemos que la razón de ser de la dinámica del sistema autoevidenciable es minimizar la energía libre.

Una vez expuestas estas relaciones, salta a la vista que, de hecho, hay tres maneras en las que un sistema autoevidenciable puede reducir el error de predicción y, de ese modo, minimizar la energía libre, y no solo las dos maneras obvias que he descrito anteriormente:

(1) Puede actuar (es decir, cambiar M) para alterar sensaciones (φ) de tal modo que coincidan con las predicciones del sistema. Eso es acción.

(2) Puede cambiar su representación del mundo (Qη) para producir una predicción mejor (ψ). Eso es percepción.

Y ahora, además:

(3) Puede ajustar la precisión (ω) para combinar óptimamente la amplitud de los errores de predicción (e).

Eso, sostengo yo, es conciencia.[316]

Es este proceso de optimización final, la optimización de la confianza del sistema tal como la he descrito anteriormente, el que Friston y yo asociamos con la evaluación de la energía libre en que se apoya la experiencia sentida. Las ecuaciones que formalizan estas dinámicas se incluyen en las notas.[317] Dado que la tercera ecuación es crucial, la expresaré con palabras: «La tasa de cambio de precisión (ω) a lo largo del tiempo depende de la cantidad de energía libre (F) que cambie cuando se cambia la precisión. Esto significa que la precisión parecerá estar tratando de minimizar la energía libre.[318] La tasa de este proceso de minimización de energía libre es la diferencia entre la varianza (la precisión inversa) y la suma de los errores de predicción al cuadrado (e.e)».[319] En términos más básicos, la tercera ecuación cuantifica el modo en que el ajuste de precisión en curso implementa la ley de Friston, junto con la acción y la percepción. En pocas palabras: cuantifica la manera en que la conciencia contribuye a la acción, a la percepción y a la actualización del modelo (y, con ello, a minimizar la energía libre).

La figura 17 permite visualizar estas dinámicas.

Conceptualmente, la precisión es un determinante clave de la minimización de la energía libre y, por lo tanto, de la activación de los errores de predicción. La precisión determina qué errores de predicción se seleccionan (y, por ende, en última instancia, cómo representamos el mundo y nuestras acciones sobre él). Si precisión es «excitación» (y lo es), eso explica, tanto desde el punto de vista formal

como mecanicista, por qué la optimización de la confianza es siempre y solamente un proceso endógeno. La conciencia debe venir de dentro. [320]

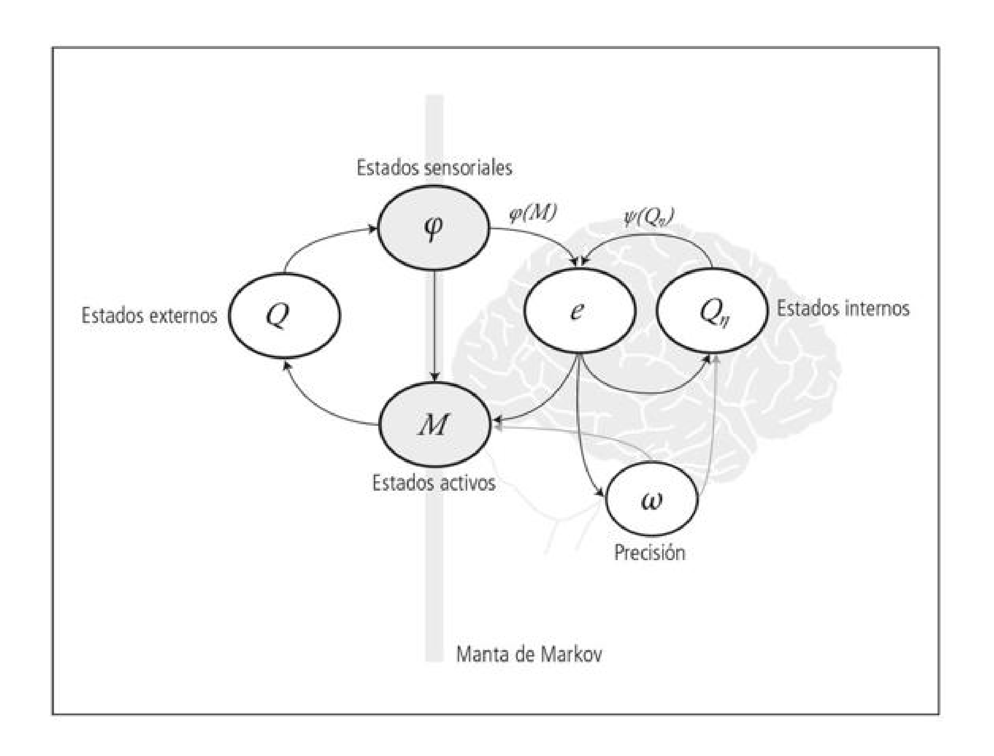


Figura 17

. Dinámica de un sistema autoevidenciable equipado con optimización de la precisión. Los símbolos se expresan verbalmente en el texto. (Q indica la realidad externa propiamente dicha, que está oculta al sistema y, en consecuencia, no aparece en las ecuaciones).

El proceso que he introducido aquí se manifiesta de muchas maneras. Considerémoslo una vez más en términos biológicos (es decir, fisiológicos y psicológicos). En el ámbito exteroceptivo, se manifiesta como atención y atenuación, asociado con el aumento y la disminución de la percepción sensorial.[321] En el ámbito propioceptivo, se corresponde con la precisión de los ofrecimientos motores (usos posibles de objetos), de la clase asociada a la selección y realización de metas.[322] En el ámbito interoceptivo determina literalmente las «corazonadas», esto es, la mejor explicación de las

señales interoceptivas que han sido activadas o excitadas.[323] Sin embargo, es muy importante señalar que todos estos aspectos (exterocepción, propiocepción e interocepción) pueden darse sin conciencia; la conciencia es la sensación de esos aspectos.

Para resumir, digamos que la tarea de la precisión es excitar representaciones (y expectativas). En ausencia de precisión, los errores de predicción no logran inducir ninguna síntesis perceptiva ni conducta motivada. Dicho de otro modo, sin precisión los errores de predicción se verían secuestrados en el momento de su formación. Eso es lo que ocurría en los pacientes de mutismo acinético de Oliver Sacks, por ejemplo.

Esta formulación de la precisión implica los mecanismos neuromoduladores descritos en el capítulo 6. Estos mecanismos generan estados alterados de conciencia[324] y sueños,[325] y son el blanco de las drogas que alteran la conciencia (psicotrópicas y psicodélicas).[326] Esta formulación de la precisión también confiere cierta validez a las versiones neuromodeladoras de la teoría del «espacio de trabajo global» (analizada brevemente en el capítulo 4).[327]

No ha de sorprender que el papel de la precisión en la psicopatología sea un tema importante en el floreciente campo de la psiquiatría computacional.[328] Recordemos el caso del señor S., en el que veíamos lo que ocurre cuando se les asigna demasiado poco peso a las señales de error. La figura 17 nos permite ver el modo en que su triángulo decisorio y, por tanto, el sistema reticular activador (ω) asignaron demasiado peso a su modelo predictivo (Qη) y demasiado poco a sus errores de predicción (e). Como han visto, se trata de algo que estaba totalmente relacionado con los sentimientos del señor S.

Sé que suena raro hablar de la conciencia en términos mecanicistas como estos, y es así porque he estado describiendo las leyes subyacentes a la experiencia fenoménica más que dicha experiencia en sí. Al exponer esas leyes he intentado mostrar que la conciencia es parte de la naturaleza, que no existe en un universo paralelo y que no es algo que está fuera del alcance de la ciencia.

Ahora quiero pedirles que crucen conmigo un Rubicón.

El presente capítulo aborda dos cuestiones: por qué y cómo surge la conciencia, no biológicamente, sino desde los puntos de vista formal y

mecanicista. Más en concreto, pregunta: (1) ¿por qué y cómo el proceso biológico de priorización de necesidades que he descrito antes se relaciona con las leyes de la minimización de la energía libre?; (2) ¿por qué y cómo este proceso mecanicista provoca que algunos sistemas autoorganizados se sientan como algo? La cuestión que he estado abordando en las últimas páginas, desde la perspectiva de la física estadística, es la primera; por lo tanto, ahora debo explicar con sencillez por qué y cómo las dinámicas estadístico-mecánicas que he descrito generan experiencias sentidas. ¿Cómo puede tener sentimientos un simple sistema de procesamiento de la información? Para explicarlo, tengo que pedirles que den un salto que muchos científicos naturales son reacios a dar (muy en detrimento de la ciencia y, sobre todo, de la ciencia mental). Se trata de analizar la dinámica mecanicista que he descrito desde el punto de vista del sistema. Les pido que reemplacen la perspectiva objetiva de tercera persona que hasta ahora hemos adoptado con la dinámica en el presente capítulo por una perspectiva de primera persona, la perspectiva subjetiva del sistema autoevidenciable propiamente dicho. Estoy pidiéndoles que adopten el punto de vista del sistema para empatizar con él.[329]

Dos hechos que ya he explicado justifican este salto. El primero es que la experiencia sentida solo puede registrarse desde la perspectiva subjetiva. Por lo tanto, descartar la perspectiva subjetiva significa excluir de la ciencia el rasgo más esencial de la mente. Eso fue lo que hicieron los conductistas, sentando las bases para medio siglo en que la neurociencia no consiguió lidiar adecuadamente con la conciencia. El segundo hecho es el siguiente: he demostrado en términos formales y mecanicistas cómo surge la subjetividad de los sistemas autoevidenciables. Por lo tanto, adoptar la perspectiva subjetiva de un sistema autoevidenciable se justifica precisamente por el hecho de que posee mismidad.

Para aclarar este punto, diré que la energía libre y sus precisiones constituyentes solo son experimentables dentro de un sistema cuando este se concibe subjetivamente, desde el punto de vista del sistema; las experiencias no se pueden observar como experiencias desde fuera, objetivamente.

Asimismo, he explicado en términos causales y observantes de la ley por qué y cómo la mismidad de tales sistemas es intencional. Los sistemas autoorganizados con la dinámica que he descrito tienen un objetivo y un propósito, a saber: sobrevivir. Eso significa que poseen un sistema de valores, el mismo que sustenta todas las vidas.

La intencionalidad de los sistemas autoorganizados dinámicos los obliga a formular preguntas acerca de sus propios estados en relación con las perturbaciones entrópicas que los rodean, y eso es lo que los convierte en sistemas «autoevidenciables». Siempre deben preguntar: «¿Qué le ocurrirá a mi energía libre si hago esto?». Por otra parte, los sistemas autoevidenciables complejos (como nosotros, los vertebrados) tienen que hacer esa pregunta en relación con múltiples variables categóricas; así, las respuestas —nuestras estadísticas vitales— han de estar tanto cuantificadas como cualificadas. Por último, deben modular su nivel de confianza en las respuestas que reciben.

Lo que describo aquí en términos abstractos y técnicos no es nada demasiado complicado. Es algo que sabemos por experiencia personal. Lo que experimentamos todo el tiempo son breves avisos de sensaciones en respuesta a nuestro movimiento a través del mundo a medida que comprobamos que todo está como esperábamos encontrarlo (y a medida que intentamos cerrar la brecha de una manera u otra cuando no lo está). ¿O no podríamos resumir así lo que es la experiencia para ustedes y para mí?

Si combinamos todos estos hechos acerca de la subjetividad y la intencionalidad de los sistemas autoevidenciables complejos, llegamos a la siguiente conclusión: las respuestas suscitadas por el equipo (en el sentido de Wheeler) que fluyen subjetivamente a partir de los tipos de preguntas que los sistemas como nosotros estamos obligados a formularnos deben poseer valor existencial y múltiples cualidades. Nuestra confianza en esas respuestas variables (los «fenómenos») que registramos debe ser subjetiva, valenciada y cualificada.

Y experimentar conscientemente es eso y no otra cosa. Las respuestas suscitadas por el equipo, al menos en nuestro caso, el de los vertebrados —y sin duda también en otros organismos—, son sentidas.

Para ayudarles a cruzar el Rubicón, por favor recuerden que los sentimientos evolucionaron. El amanecer de la conciencia dio lugar a fenómenos muy simples, como sentir demasiado calor. La evidencia sensorial cada vez más precisa que predice la desaparición de un sistema autoevidenciable que se sobrecalienta se siente simplemente como un «hace mucho calor» dirigido al sistema. Hicieron falta eones para que formas tan elementales de afecto se elaboraran a lo largo de una profunda jerarquía predictiva y, en última instancia, produjeran el «mundo» de Merker, «tridimensional, panorámico y completamente articulado, compuesto de objetos sólidos con forma: el mundo de nuestra experiencia fenoménica familiar».

En un mundo así sentimos cómo es ser un sistema con las dinámicas que he descrito. Los sentimientos son estados subjetivos variables, valorados existencialmente, con cualidades diferenciadas y grados de confianza. Esa es la materia de la conciencia. Ahora comprendemos por qué tiene que ser así.

[301] Rolls, 2019, p. 10; la cursiva es mía. Véase también mi comentario: Solms, 2019b.

[302] «No deberían utilizarse más cosas de las necesarias». Si hay varias formas posibles de poder explicar algo, la correcta será probablemente la que recurra a menos conjeturas.

[303] Roger Sessions, New York Times, 8 de enero de 1950.

[304] Friston (2009, p. 299) define el papel de la precisión como el de «controlar la influencia relativa de las expectativas previas a varios niveles». Como señala Hohwy (2013, p. 199): «Si se aumenta la ganancia [precisión] de una señal, hay que reducir la ganancia de las demás. De lo contrario, la noción de ganancia pierde el sentido: las ponderaciones deben sumar uno. Así, si las expectativas de precisión hacen aumentar la ganancia para un error de predicción, disminuirá la ganancia para los demás».

Cf. Clark, 2016, p. 313: «Feldman y Friston (2010) señalan que la precisión se comporta como si ella misma fuera un recurso limitado, en tanto en cuanto hacer aumentar la precisión para determinadas unidades de error de predicción obliga a reducir la de otras. También comentan, de forma algo intrigante (op. cit., p. 11): "El motivo de que la precisión se comporte como un recurso es que el modelo generativo contiene creencias previas de que la precisión logarítmica se redistribuye en los canales sensoriales de un modo sensible al contexto pero se conserva en todos los canales"». Esto nos remite a lo que especulaba en el capítulo 6 acerca de que cada neuromodulador ajusta distintos aspectos de la excitación. Curiosamente, las oscilaciones gamma (error) responden a la acetilcolina.

[305] Los neurofisiólogos también llaman «ganancia» a la precisión. En este punto, la variedad terminológica empleada por los científicos de las distintas disciplinas puede llevar a bastante confusión. La «actividad» sináptica (= neurotransmisión) está modulada por la

«ganancia» postsináptica (= neuromodulación), que con el tiempo determina la «eficiencia» (= neuroplasticidad). Estas mismas variables también son descritas por los científicos computacionales como «estados» de la señal, «precisiones» de la señal y «parámetros» de la señal, respectivamente. En términos generales, la neurotransmisión implica estados, la neuromodulación implica precisiones y la plasticidad implica parámetros.

[306] Friston lo llama «búsqueda epistémica».

[307] En este párrafo parafraseo estrechamente a Clark, 2015, p. 70.

[308] Esto da lugar a la «ceguera por falta de atención», cuyo ejemplo más conocido es el que se ve aquí: https://www.youtube.com/watch?v=vJG698U2Mvo (véase Chabris y Simons, 2010).

[309] Varela, Thompson y Rosch, 1991, p. 198. Varela llama a este enfoque de la percepción «enactivo». Véase Clark, 2015, p. 173: «Lo que más preocupa de un enfoque enactivo de la percepción no es determinar cómo se va a recuperar un mundo independiente del perceptor, sino determinar los principios comunes o los vínculos legítimos entre los sistemas sensoriales y los sistemas motores que explican cómo se puede guiar perceptivamente la acción en un mundo dependiente del perceptor».

[310] Es numerosa la literatura experimental que apoya empíricamente esta conceptualización formal de la precisión. Baste con mencionar por ahora un estudio que ilustra la cuestión básica. Mediante IRMf, Hesselmann *et al.* (2010) analizaron la actividad cerebral a nivel base en condiciones que manipulaban las expectativas de precisión. Para ello presentaron a los participantes del estudio estímulos (visuales y auditivos) con distintos niveles de ruido; cuando los participantes esperaban precisión, se hacía aumentar la ganancia, y cuando esperaban imprecisión, se reducía. Es decir, se facilitaba la predicción descendente cuando se esperaban señales sensoriales imprecisas. Véase Feldman y Friston, 2010, para una explicación completa de cómo la optimización de la precisión explica la atención endógena y exógena.

[311] Como recordarán de la p. 48, en una carta a Fliess fechada a 20 de octubre de 1895, Freud escribió lo siguiente: «En el transcurso de una noche ajetreada […] se levantaron de repente las barreras, cayeron los velos y fue posible ver desde los detalles de las neurosis hasta los determinantes de la conciencia. Todo parecía encajar, los engranajes estaban bien colocados; daba la impresión de que era

realmente una máquina y que pronto funcionaría sola». Y luego continuó: «Los tres sistemas de neuronas [φ, ψ y ω], los estados libres y ligados de la cantidad [Qη], los procesos primario y secundario, la tendencia principal y la tendencia de compromiso del sistema nervioso, las dos reglas biológicas de la atención y la defensa, los signos de cualidad [ω], realidad y pensamiento, el estado de los grupos psicosexuales, el condicionamiento sexual de la represión y, finalmente, los determinantes de la conciencia como función perceptiva…, ¡todo encajaba y sigue encajando! Lógicamente, no puedo contenerme de gozo». Sin embargo, poco tiempo después se dio cuenta de que los engranajes no estaban tan bien colocados ni todo encajaba tanto. Y decidió abandonar el proyecto. Cuando Friston y yo lo resucitamos —y confío en que, aprovechando más de un siglo de avances de la neurociencia, lo completamos (al menos un esbozo; véase Solms, 2020b)—, recordé las conmovedoras palabras de Freud: «daba la impresión de que era realmente una máquina y que pronto funcionaría sola». Véase el capítulo 12.

[312] Nótese que ω no es un vector, sino un escalar (escala una matriz de precisión). Nótese también que a los términos Q y Qη se les dio un uso algo distinto en Solms y Friston, 2018, al que les doy aquí.

[313] Donde e es un vector y F es un escalar. La notación de punto en las ecuaciones siguientes implica un producto punto (es decir, una multiplicación matricial o vectorial). Estas ecuaciones son íntegramente atribuibles a la aportación de Friston a nuestra colaboración.

[314] $F \approx -\log P(\varphi(M))$.

[315] $E[F] \approx E[-\log P(\varphi)] = H[P(\varphi)] = -\frac{1}{2} \cdot \log(|\omega|)$, donde $E[\cdot]$ denota expectativa o promedio, y P y H denotan probabilidad y entropía, respectivamente (como antes); bajo supuestos gaussianos sobre fluctuación aleatoria.

[316] La importancia general de la optimización de la precisión se reconoció hace tiempo, especialmente en relación con la atención (véanse Friston, 2009; Feldman y Friston, 2010). Sin embargo, la primera científica que reconoció su importancia para la propia conciencia fue Katerina Fotopoulou (2013, p. 35): «Un aspecto central de la conciencia puede servir para registrar la mencionada cualidad de "incertidumbre" y su cualidad opuesta, la precisión. Esta perspectiva choca con la visión (intuitiva y antigua) del núcleo de la conciencia afectiva como cualidad hedónica de control, expresada por Solms en términos freudianos como la serie placer-displacer. En su lugar,

propongo que la cualidad central de este aspecto de la conciencia (a diferencia de la conciencia perceptiva [...]) es una especie de certeza-incertidumbre, o principio de desambiguación». Debo observar aquí que, en lo relativo a la serie placer-displacer frente a certeza-incertidumbre, creo que Fotopoulou me malinterpretó ligeramente, como señalé en mi respuesta: «No estoy [...] seguro de lo que quiere decir cuando afirma que la incertidumbre la controla el afecto en lugar de la calidad hedónica. A mi modo de ver, la calidad hedónica es precisamente nuestra medida de la incertidumbre» (Solms, 2013, p. 81).

[317] $(\partial/\partial t)\, M = -(\partial F/\partial M) = -(\partial F/\partial e)\,(\partial e/\partial M) = (\partial \varphi/\partial M) \cdot \omega \cdot e$ (1)

$(\partial/\partial t)\, Q\eta = -(\partial F/\partial Q\eta) = -(\partial F/\partial e)\,(\partial e/\partial Q\eta) = -(\partial \psi/\partial Q\eta) \cdot \omega \cdot e$ (2)

$(\partial/\partial t) * \omega = -(\partial F/\partial \omega) = (1/2) \cdot (\omega - 1 - e \cdot e)$ (3)

Donde ∂ denota una derivada parcial y t denota el tiempo, y el error de predicción y la energía libre son:

$$e = \varphi(M) - \psi(Q\eta)$$

$$F = (1/2) \cdot (e \cdot \omega \cdot e - \log(\omega))$$

Presento estas ecuaciones a grandes rasgos, como caricaturas de alto nivel. Necesitan un mayor desarrollo, parte del cual solo podrá hacerse cuando se pongan en práctica (véase el capítulo 12). Por ejemplo, como ya se ha dicho, en tratamientos más completos también habría que considerar modelos generativos jerárquicos (con precisiones en cada nivel) y dar cabida a la incertidumbre condicional sobre los estados externos. Además, las ecuaciones agrupan todos los errores de predicción sensorial, incluidas las modalidades exteroceptiva, propioceptiva e interoceptiva. Obsérvese que el término propioceptivo es aquí sinónimo de cinestésico (Friston utiliza propioceptivo simplemente por armonía aliterativa con exteroceptivo e interoceptivo).

[318] Técnicamente, esto se denomina «descenso por gradiente»,

donde el gradiente es la tasa de cambio de la energía libre con precisión.

[319] Bajo nuestros supuestos simplificadores sobre la codificación de las creencias bayesianas (véase anteriormente).

[320] En este sentido, puede decirse que la precisión desempeña el papel del «demonio de Maxwell», un experimento mental creado por el físico James Clerk Maxwell (1872): un demonio controla una pequeña puerta entre dos cámaras de gas. Las moléculas de gas flotan a distintas velocidades. Cuando las moléculas más rápidas de la primera cámara alcanzan la puerta, el demonio la abre y la cierra, muy brevemente, para que pasen a la segunda cámara, mientras que las más lentas se quedan atrás. Como las moléculas más rápidas generan más calor que las más lentas, esto disminuirá la entropía, algo que no puede ocurrir sin trabajo. Si equiparamos el paso de moléculas en esta analogía con la neurotransmisión de señales sensoriales, entonces la ponderación de precisión (neuromodulación) hace lo mismo que el demonio de Maxwell: selecciona señales sensoriales para confundir la segunda ley de la termodinámica. Obsérvese que, en términos de la dinámica de sistemas aquí descrita, la conciencia es la actividad del demonio de Maxwell en sí mismo; no es el paso de moléculas que dicha actividad permite. Esto significa que la conciencia es la optimización de la precisión con respecto a la energía libre; no es el paso de mensajes a través de una jerarquía predictiva. En la figura 17, por tanto, la conciencia es la actividad de ω (precisión), que determina la influencia relativa de e (señales de error) sobre Qη (el modelo interno). El demonio de la precisión modula la influencia de los errores en relación con el modelo. La conciencia se basa en un atributo de las creencias en contraposición al contenido de las creencias (es decir, se basa en la precisión variable de [o la confianza en] las creencias sobre los estados de cosas internos y externos). La actividad de este demonio provoca secuelas sensoriales a través de la amplificación o atenuación de los errores de predicción; la optimización de la precisión no es inherente a las propias señales de error. Para más información sobre las implicaciones biológicas del demonio de Maxwell, véase el excelente libro de Paul Davies (2019) que no citamos en su momento porque, lamentablemente, apareció después de nuestra propia aplicación del concepto a la conciencia (Solms y Friston, 2018).

[321] Brown et al., 2013; Feldman y Friston, 2010; Frith, Blakemore y Wolpert, 2000.

[322] Cisek y Kalaska, 2010; Frank, 2005; Friston et al., 2012; Friston,

Schwartenbeck, FitzGerald et al., 2014; Moustafa, Sherman y Frank, 2008.

[323] Hohwy, 2013; Seth, 2013; Ainley et al., 2016. Sobre sensibilidad interoceptiva y modulación social del dolor, véanse también Crucianelli et al., 2017; Fotopoulou y Tsakiris, 2017; Krahe et al., 2013; Decety y Fotopoulou, 2015; Paloyelis et al., 2016; y Von Mohr y Fotopoulou, 2017.

[324] Ferrarelli y Tononi, 2011; Lisman y Buzsaki, 2008; Uhlhaas y Singer, 2010.

[325] Véanse Hobson, 2009; y Hobson y Friston, 2012, 2014. Aunque, en mi opinión, una explicación satisfactoria de los sueños en este marco debe partir de su carácter consciente (y afectivo).

[326] Nour y Carhart-Harris, 2017.

[327] Dehaene y Changeux, 2011; Friston, Breakspear y Deco, 2012.

[328] Montague et al., 2012; Corlett y Fletcher, 2014; Friston, Stephan, Montague y Dolan, 2014; Wang y Krystal, 2014.

[329] Para un análisis sobre «empatía», véase Solms, en prensa. El término original en alemán es Einfühlung, que literalmente significa «sentir en».

10

De vuelta a la corteza cerebral

Como hemos visto repetidas veces a lo largo de nuestro viaje, la falacia cortical plantea muchos interrogantes: si los pioneros de las neurociencias conductuales no hubieran estado tan impresionados por la gran extensión de nuestra corteza cerebral, o no hubieran estado tan imbuidos de la idea filosófica de que la vida mental surge de la asociación con imágenes de la memoria, podríamos haber descubierto la verdadera fuente de la conciencia mucho antes. Resulta irónico y perturbador para la historia de la ciencia mental que hace más de un siglo Freud poseyera ya tantas piezas del rompecabezas. Tenía las pistas tanto neurológicas como psicológicas ante los ojos, pero, en lo que respecta a la conciencia, incluso él cayó presa de nuestra fijación colectiva con la corteza cerebral, una obsesión cuyo coste, por si lo hemos olvidado, puede medirse en algo más que tiempo perdido.

Sin embargo, a pesar de todo lo anterior, es innegable que la corteza desempeña un papel fundamental y que nuestra experiencia cotidiana está íntimamente ligada a la dinámica del procesamiento cortical. En este capítulo volveremos a ocuparnos de este mal interpretado piso superior del cerebro para ver qué aporta a nuestra explicación de la conciencia. Como veremos, el carácter de muchos de los rasgos más comunes de nuestra experiencia cotidiana se deriva de lo que hace la corteza, pero no como creíamos antes.

Lo más obvio en este sentido es que el mundo, tal y como lo experimentamos, se genera literalmente a partir de representaciones corticales. En el marco de la codificación predictiva, por extraño que parezca, lo que percibimos es una realidad virtual fabricada a partir de los propios materiales de construcción de la mente.

Se trata de una noción radical, si la comparamos con el punto de vista del sentido común, pero la idea de que la experiencia perceptiva es algo autogenerado está ampliamente aceptada en la neurociencia contemporánea. Tomemos, por ejemplo, lo que dijo Semir Zeki ya en 1993 sobre la visión del color. Escribió que el color es «una propiedad del cerebro, una propiedad que se traslada a las superficies exteriores, una interpretación que dota de determinadas propiedades físicas a los objetos».[330] Y prosigue:

Supongamos que se observa un área iluminada aislada con luz de onda larga [...]. El área produce un registro de luminosidad elevado para luz de cualquier onda, ya que la única comparación que el cerebro puede hacer en estas condiciones es entre la luz reflejada del área iluminada y el entorno oscuro. Así, la luz de onda larga produce una luminosidad elevada, mientras que la luz de onda media y corta, al estar ausente, no produce luminosidad alguna. El sistema nervioso asigna así el color rojo al área. [331]

Quiero llamar la atención sobre las palabras que elige Zeki: el cerebro asigna el color rojo al mundo en general tras preguntarse por las intensidades relativas y las longitudes de onda de luz. Pinta el mundo con números. Lo mismo ocurre con las propiedades fenoménicas que caracterizan nuestras otras modalidades de percepción: sonidos, sabores, sensaciones somáticas y olores. El cerebro asigna estas cualidades al mundo.

Sospecho que a muchos lectores les cuesta creer que lo que están viendo ahora mismo no es simplemente «lo que está ahí». Puedo imaginar cómo se preguntan: «¿De dónde viene mi percepción de estas palabras en la página?». Quizá les ayude considerar que lo que están viendo en este momento se parece muy poco a las señales sensoriales que están recibiendo. Estas señales comienzan como ondas de luz que impactan en la retina. Sus células fotosensibles (los bastones y conos) responden a las ondas luminosas generando impulsos nerviosos. Esos impulsos —no las ondas luminosas propiamente dichas— se propagan a través de los nervios ópticos hasta la corteza cerebral en forma de trenes de impulsos nerviosos (véase fig. 11). ¿Por qué percibimos estos trenes de impulsos nerviosos —001111101101— como imágenes en movimiento en el mundo?

Las neuronas del cuerpo geniculado lateral y de la corteza occipital «de proyección» que responden a los impulsos retinianos están dispuestas topológicamente, lo que permite crear imágenes mediante el mapeo de las superficies retinianas (véase fig. 6), pero hay que tener en cuenta que no hay bastones ni conos cerca del centro de la retina, donde nace el nervio óptico. Por lo tanto, debería haber un agujero negro cerca del centro del campo visual. ¿Cómo desaparece el agujero? La respuesta es que, a partir del contexto y de la memoria, inferimos lo que debería haber en el «punto ciego» y luego rellenamos los huecos.[332]

Debería haber dicho que hay un agujero negro en los campos visuales,

en plural, porque no debemos olvidar que tenemos dos. Esto nos lleva a otra pregunta: ¿por qué no vemos dos imágenes? No me refiero al hecho de tener dos ojos, es fácil imaginar cómo pueden superponerse dos mapeos casi idénticos, pero no es eso lo que ocurre.[333] Lo que ocurre es que las células de las mitades izquierdas de ambas retinas se proyectan en el lóbulo occipital derecho, mientras que las mitades derechas se proyectan en el lóbulo izquierdo. Esto quiere decir que lo que tenemos realmente en la corteza visual son dos representaciones diferentes de las superficies retinianas (una de la mitad izquierda de este libro y otra de la mitad derecha),[334] con un abismo anatómico entre ellas: la fisura longitudinal que divide los hemisferios cerebrales. ¿Cómo se convierten los dos campos en la imagen unificada que vemos? (Es cierto que están coordinadas por axones a través del cuerpo calloso, pero las personas a las que les falta el cuerpo calloso también ven una sola imagen).[335] Además, hay que tener en cuenta que los campos visuales (tal y como están representados en los lóbulos occipitales) están boca abajo y al revés en relación con las imágenes que vemos. Si añadimos que los ojos se mueven haciendo un barrido a una velocidad de unas tres veces por segundo, por no hablar del movimiento constante de la cabeza, ¿cómo percibimos una escena visual estable y correctamente orientada?

Lo que quiero decir (que hay pocas similitudes entre lo que vemos y los aportes sensoriales que llegan a la corteza cerebral) se ilustra claramente en pacientes neurológicos que tienen lesionados los mecanismos por los que normalmente convertimos lo que llega a la corteza en lo que vemos. Hace muchos años presenté un caso de este tipo: un niño de doce años (W. B.) con abscesos bilaterales en los lóbulos frontales, causados por una sinusitis galopante.[336] Periódicamente veía el mundo girado ciento ochenta grados. Los síntomas y signos de este paciente eran idénticos a los descritos en veintiún informes de casos anteriores que encontré dispersos por la literatura científica mundial, desde 1805, lo que da credibilidad a su descripción subjetiva (véase fig. 18 para algunas pruebas objetivas).

Actualmente tengo una paciente aún más interesante que está siendo investigada por mi alumna de doctorado Aimee Dollman. No puedo informar de todos los detalles, porque el caso sigue en estudio y todavía no se han publicado las conclusiones. La paciente es una joven muy inteligente con disgenesia cortical (anatomía anormal) de los lóbulos occipitales, que representa el mundo casi exactamente de la forma en la que acabo de decir que no lo experimentamos: tal y como está dispuesto anatómicamente en la corteza visual. Esto ocurre en especial cuando utiliza su memoria visual (es decir, su modelo predictivo). Ve dos campos separados, boca abajo y al revés (no

siempre de forma simultánea). Su modelo visual del mundo no lleva a cabo las inferencias correctoras habituales con las que orientamos e integramos los campos visuales. Por tanto, sus predicciones son erróneas y su experiencia visual no cuadra con la de sus otras modalidades sensoriales. En consecuencia, a veces se confunde sobre la dirección en la que se mueve su cuerpo (especialmente cuando viaja en trenes y aviones) y comete errores burdos al moverse por su entorno. Además, tiene dificultades para inferir objetos invariantes a partir de datos visuales fugaces (como la ortografía de las palabras en el ojo de su mente cuando la tiene que abstraer a partir de escrituras y tipos de letra variables, o la identidad de caras abstraídas en condiciones de iluminación fluctuante y diferentes ángulos de visión).

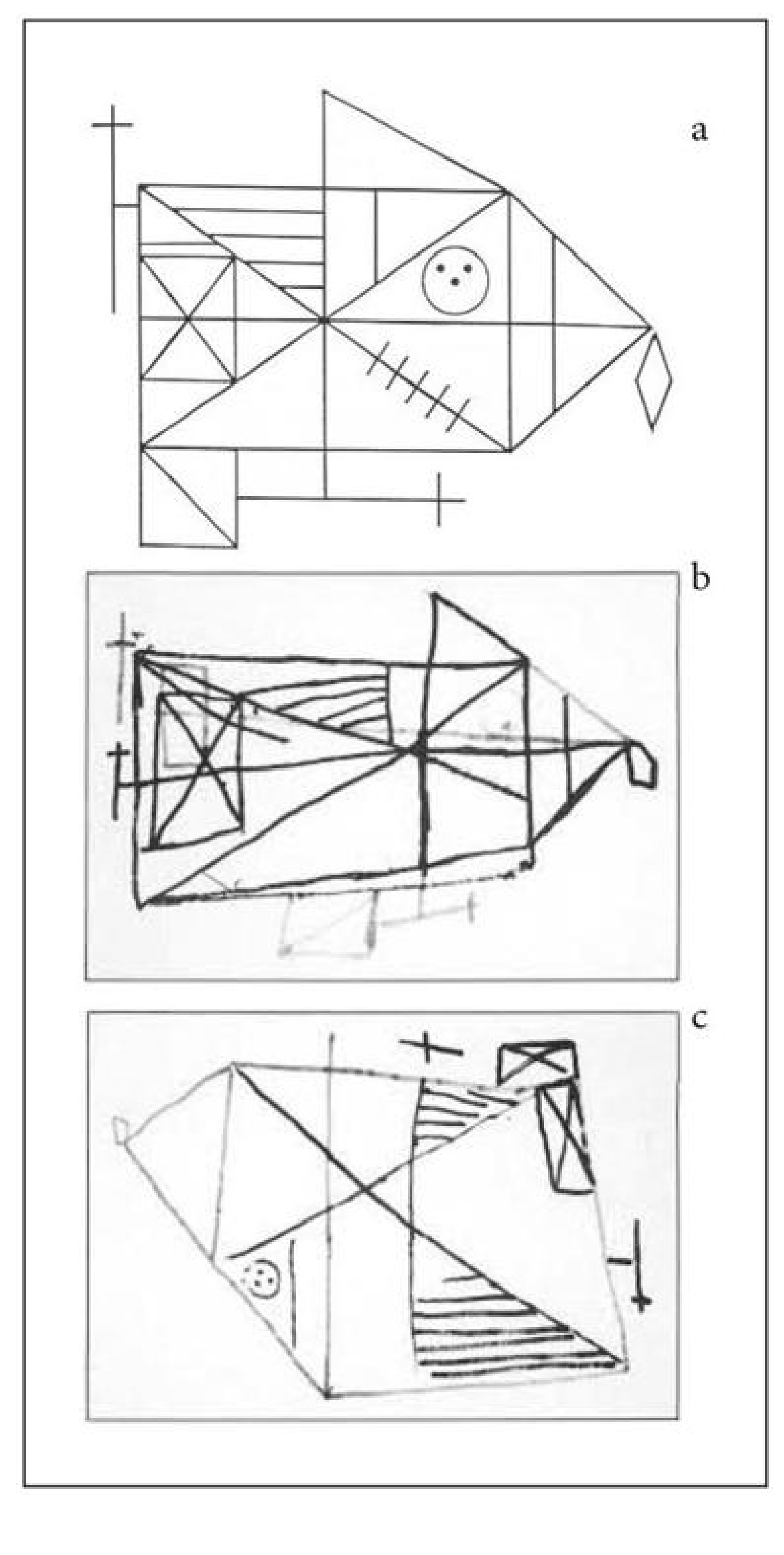

a
b
c

Figura 18

. (a) = la figura compleja de Rey mostrada al paciente W. B.; (b) = la misma figura copiada por él; (c) = la figura dibujada por él de memoria. Estos dibujos proporcionan pruebas objetivas de la inversión del modelo predictivo del mundo de W. B. La figura compleja de Rey es difícil de reproducir en el mejor de los casos. ¿Por qué este paciente gravemente enfermo aumentaría la dificultad dibujándola al revés?

En otras palabras, la corteza visual de la paciente recibe el tipo de información que recibimos todos, pero no puede generalizar a partir de las señales sensoriales confusas e inferir automáticamente los objetos estables que representan (por ejemplo, reconocer una cara familiar). Ha sufrido estas anomalías toda su vida y ha desarrollado procedimientos elaborados de compensación. Dado que su corteza visual de «asociación» no integra automáticamente los dos campos visuales ni da vuelta a la escena, la paciente ajusta su representación del mundo haciendo inferencias deliberadas. Por ejemplo, cuando le pedí que identificara la ubicación de una ciudad conocida en un mapa mudo, me preguntó: «¿Debo mostrar dónde percibo que está o dónde sé que está?». Cuando «percibe» que algo está situado al oeste, «sabe» que debe estar situado al este.

Sin embargo, en lugar de acudir a estos raros trastornos neuropsicológicos, voy a ilustrar la naturaleza autogenerada de la percepción mediante el fenómeno que se utiliza convencionalmente para este fin, a saber: la rivalidad binocular.

Este fenómeno se describió por primera vez en 1593 y ha ocupado un lugar destacado en la obra seminal de Helmholtz sobre el tema de la inferencia consciente.[337] Consiste en la presentación simultánea de imágenes diferentes a cada ojo utilizando un estereoscopio de espejos. Por ejemplo, se nos presenta una cara al ojo izquierdo y un edificio al derecho. En estas condiciones artificiales, la experiencia visual se desarrolla de forma «biestable», por lo que no vemos una mezcla superpuesta de ambas imágenes, sino una alternancia entre ellas. Vemos un edificio y luego una cara, y luego un edificio y luego una cara, en lugar de una combinación de edificio y cara. Esto demuestra claramente la distinción entre la señal objetiva que se transmite al cerebro y el percepto subjetivo que genera. Helmholtz concluye: «En estos casos, la interpretación [de la señal visual] vacila de tal manera que el observador tiene diferentes experiencias que se presentan de forma sucesiva para una imagen retiniana que no cambia».[338] Por

lo tanto, al igual que con la visión del color, lo que se experimenta es una inferencia sobre el aporte sensorial, no el aporte sensorial en sí.

En la vida cotidiana podemos tener experiencias muy similares, como cuando «vi» a mi amiga británica Teresa en el aeropuerto de Ciudad del Cabo. Todas estas ilusiones demuestran que lo que percibimos está generado en gran medida por las expectativas que tengamos. En términos bayesianos, se considera que la rivalidad binocular demuestra que si la hipótesis previa que mejor se ajusta a los datos sensoriales (la alta probabilidad de que se esté viendo una cara-edificio) no cuadra con los conocimientos previos (la baja probabilidad de que existan caras-edificio), entonces rechazamos la hipótesis. La inferencia de que estamos viendo un edificio supera a la inferencia de que estamos viendo una cara-edificio y, por lo tanto, lo que experimentamos es un edificio. Sin embargo, cuando sometemos a prueba esta hipótesis posterior (como una nueva hipótesis previa) solo se ajusta a la mitad de las pruebas sensoriales. Los conocimientos previos dictan que una cara es tan probable como un edificio. Entonces, cambiamos de opinión y deducimos que se trata de una cara, y es lo que experimentamos. A continuación, sometemos a prueba esta nueva hipótesis y, de nuevo, solo se ajusta a la mitad de las pruebas sensoriales. Y así sucesivamente…

La interpretación bayesiana de la rivalidad binocular está ampliamente aceptada. Lo que me convence es el hecho de que cuando se muestran dos imágenes que, combinadas, tienen una alta probabilidad a priori, entonces se percibe una imagen combinada. Por ejemplo, las imágenes presentadas dicópticamente de un canario y una jaula se perciben fácilmente como un canario dentro de una jaula.

Lo que percibimos no es lo mismo que nos llega a través de los sentidos. Lo que percibimos es una inferencia. Y los materiales de los que se deriva esa inferencia son, en su mayor parte, nuestro modelo de predicción cortical derivado de experiencias pasadas (es decir, previsibles).[339]

Eso nos dice algo sobre lo que la corteza cerebral aporta a la conciencia. Pero ¿y al revés? ¿Qué es lo que la conciencia aporta a la corteza cerebral?

Lo que voy a decir ahora es obvio, pero nadie más parece decirlo.[340] Se trata de lo siguiente: la conciencia cognitiva está generada por un mecanismo neuronal recientemente descubierto llamado

«reconsolidación de la memoria». Como otras tantas cosas, esta idea tiene su origen en algo que escribió Freud. Escribió que «la conciencia surge en remplazo de la huella mnémica».[341] Lo que él tenía en mente era ligeramente diferente de lo que voy a decir, porque Freud, como todos los neurólogos de su época, estaba atrapado en la falacia cortical.[342] No obstante, una vez más, es bastante sorprendente lo cerca que estaba de la verdad.

Hemos visto que los afectos plantean demandas de trabajo a la mente y que la cognición lleva a cabo ese trabajo. Para ser más exactos, la cognición consciente lleva a cabo el trabajo, porque una vez que se ha realizado y se ha restaurado la confianza en la creencia (priorizada) que se había vuelto incierta, el modelo generativo vuelve al modo automático de funcionamiento, por debajo del umbral de conciencia.[343] Aquí está, una vez más, el mecanismo de aprendizaje a través de la experiencia que he descrito repetidas veces. Esta es la razón de ser de la conciencia en la cognición. Llegamos a una situación en la que no estamos seguros de qué hacer. La conciencia acude al rescate, tanteamos el terreno sintiendo y tomamos nota de las acciones voluntarias que funcionan para nosotros. Luego, poco a poco, las lecciones que funcionan se automatizan y la conciencia deja de ser necesaria.[344]

Quiero insistir en que el trabajo cognitivo que acabo de describir ralentiza el proceso automático de actuar en el mundo. Esta es la diferencia esencial entre la acción voluntaria y la involuntaria, la cognición consciente frente a la inconsciente, la pulsión sentida frente al reflejo autónomo; la acción voluntaria tiene menos certeza y, por lo tanto, requiere más tiempo. Este proceso, que retrasa las tendencias a actuar de forma automatizada y permite retenerlas en la mente (en la memoria a corto plazo) se denomina acertadamente «memoria de trabajo». La memoria de trabajo es, literalmente, el hecho de retener en la mente los sentimientos, el afecto estabilizado y transformado en trabajo cognitivo. Como acabo de decir, si el afecto es la demanda de trabajo a la mente, entonces la cognición consciente es el trabajo mismo. Así, el afecto acompaña a la cognición y se transforma en ella. El «trabajo» en cuestión implica inhibir las tendencias de acción automática y estabilizar la intencionalidad mientras el sistema se abre camino sintiendo a través de problemas imprevistos. Esto confiere ventajas adaptativas considerables, ya que facilita soluciones viables a los numerosos problemas del mundo real que nuestros modelos generativos no pueden (todavía) predecir. Este proceso de estabilización es la función de la corteza cerebral.[345] La corteza está especializada en la incertidumbre.

Lo que todo esto implica es que el estado consciente es indeseable desde el punto de vista de un sistema autoorganizado. Volvamos al diagrama de la figura 12: la flecha exterior representa una demanda creciente de trabajo (afecto negativo) y la interior, una demanda decreciente (afecto positivo), pero el estado ideal es el punto de estabilización que representa la ausencia total de demanda. En los capítulos del 7 al 9 he planteado estas cuestiones desde una base formal mecanicista. La minimización de la energía libre es el estado ideal de los sistemas vivos, lo que significa que el nivel de sorpresa mínimo es lo ideal. Lo cual quiere decir, sencillamente, que la necesidad mínima es lo ideal. El afecto no es sino el anuncio de una necesidad saliente. Esto debería significar que sentir es algo bueno porque nos permite, como sistemas biológicos, resolver nuestras necesidades, evitando así la destrucción. Sin embargo, el estado ideal es sin duda aquel en el que todas nuestras necesidades están satisfechas automáticamente —incluso antes de sentirlas—, es decir, el estado en el que no hay incertidumbre. En este estado ideal teórico, en el que nuestras necesidades se satisfacen automáticamente, no sentimos nada. (Así es como se satisfacen la mayoría de nuestras necesidades corporales: se regulan de forma autónoma). Digo «estado ideal teórico» porque, con respecto a muchas de nuestras necesidades, sobre todo las emocionales, nunca llegamos a ese punto. Solo la pulsión de BÚSQUEDA garantiza que lo alcancemos. Esta es la mala noticia.

La buena noticia es que las señales de error con los valores de precisión más altos son las que más influyen en el modelo generativo. Como exigen el mayor cambio —porque declaran que algo se está haciendo mal—, representan mayor oportunidad de aprendizaje para que la próxima vez las cosas salgan mejor. Esto no es diferente del proceso de «aprender de los errores», que decían los padres y profesores. Ahora ya sabemos por qué no nos dejaban olvidarlos. Sin embargo, el sorprendente resultado sigue siendo que la conciencia no es un estado deseable en la cognición. Por lo tanto, a lo que todos aspiramos no es al placer (necesidad decreciente), sino a la zombificación (ausencia de necesidad). La ausencia de necesidad implica predicciones perfectas, lo que significa que no hay errores y, por tanto, no hay necesidad de aumentar la precisión de las señales entrantes, es decir, no hay sentimiento. Por fin la paz.

El aforismo de Freud «la conciencia surge en remplazo de la huella mnémica» debería tener más sentido ahora. Significa que la conciencia surge cuando un comportamiento automático conduce a un error, es decir, cuando la huella mnémica (una predicción) que produce un comportamiento no tiene el resultado esperado. Esto significa que la

predicción en cuestión debe actualizarse para tener en cuenta el error. Por lo tanto, la conciencia cortical puede describirse como «un trabajo predictivo en curso». La huella mnémica consciente está en proceso de actualización. Ya no es una huella mnémica. De ahí que surja la conciencia donde antes estaba la huella mnémica.[346]

En los últimos años del siglo XX, entendíamos la neurofisiología de este proceso en términos de «error de predicción de recompensa» (véase fig. 19, p. 271).[347] Entrados en el siglo XXI, dominamos mejor la fisiología de la actualización de la memoria, es decir, la «reconsolidación».

Los hechos básicos salieron a la luz por primera vez con el descubrimiento en la década de 1960 de que un recuerdo de miedo se puede eliminar mediante tratamiento electroconvulsivo (electrochoque), pero solo si se administra inmediatamente después de la recuperación del recuerdo.[348] Esto sugirió que el electrochoque interfiere en un proceso que devuelve los recuerdos de miedo manifiestos (activados) a su estado latente: si se administra mientras un recuerdo de miedo está activado, entonces el recuerdo se borra. Literalmente, deja de ser una huella mientras lo recuerdas. El estado activado de la memoria hace que la huella a largo plazo vuelva a ser lábil: hace que deje de ser un recuerdo.

Este fenómeno se confirmó cuando se redescubrió la reconsolidación y quedó fijada con esta denominación en 2000.[349] Si se administran inhibidores de la síntesis de las proteínas mientras está activada una huella a largo plazo, esta desaparece. (Los inhibidores de la síntesis de las proteínas impiden que se formen nuevas huellas a largo plazo). Esto se aplica también a otros tipos de memoria, no solo a la memoria del miedo. Los recuerdos a largo plazo en general se vuelven inestables cuando están en estado activado. Así es como se actualizan (y luego se vuelven a consolidar, es decir, se reconsolidan).

En las dos últimas décadas se han llevado a cabo numerosísimas investigaciones sobre la reconsolidación, que han demostrado que no solo se produce en humanos y roedores, sino también en pollos, peces, cangrejos, caracoles de agua dulce y abejas. Los estudios también han demostrado la existencia de un proceso similar a la reconsolidación en las vías de procesamiento del dolor de la médula espinal, lo que sugiere un papel muy básico para este fenómeno en el sistema nervioso central.

La memoria a largo plazo, a diferencia de la memoria a corto plazo, depende de la síntesis de nuevas proteínas, que se activa tras una

transmisión sináptica sustancial y repetitiva, que a su vez está modulada por el sistema activador reticular, es decir, por la excitación.[350] De ahí la famosa frase conocida como ley de Hebb: «Cuando las neuronas se activan juntas, las sinapsis se fortalecen». [351] Un recuerdo «activado» es un recuerdo excitado; un recuerdo excitado ya no es un recuerdo, ha pasado a estado de incertidumbre. Lo que intento transmitir aquí es que la conciencia cognitiva se reduce a una labilidad de los rastros de memoria cortical, y que esta labilidad es producto de la excitación. Seguimos llegando por diferentes caminos a la misma idea: los procesos corticales son fundamentalmente procesos inconscientes (son meros algoritmos, si los dejamos a su aire). La conciencia (toda ella) procede del tronco encefálico.[352]

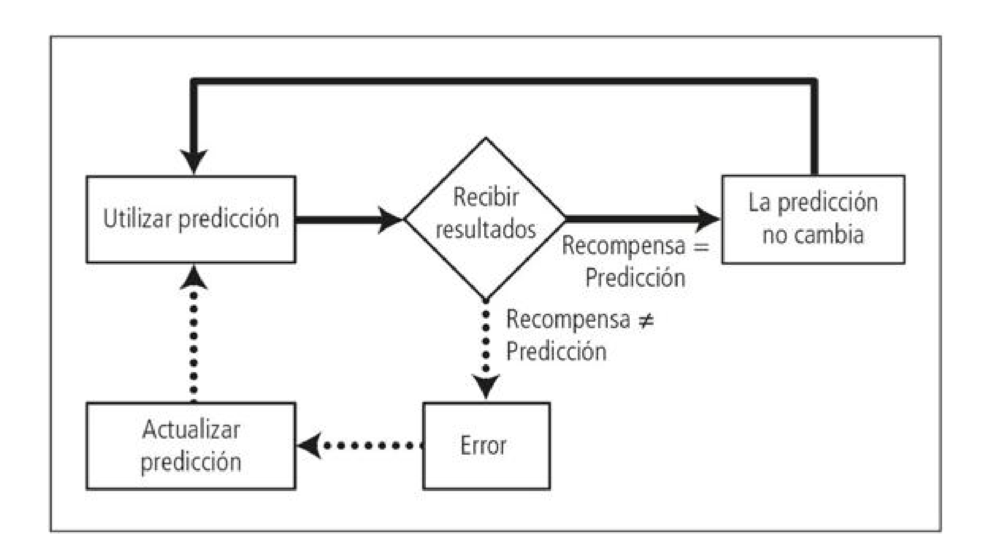


Figura 19

. Esquema de aprendizaje por error de predicción de recompensa. La secuencia comienza con la casilla «Utilizar predicción», que lleva a dos resultados posibles, a la derecha de la casilla. O bien hay un error de predicción (cuando el resultado difiere de la predicción), lo que nos lleva a la casilla «actualizar predicción», o bien no hay un error de predicción (cuando el resultado coincide con la predicción), lo que nos lleva a la casilla «la predicción no cambia», de modo que el comportamiento permanece inalterado. La etapa de «actualizar predicción» corresponde a lo que ahora llamamos reconsolidación.

Aun así, es como si nos esforzáramos por alcanzar la zombificación. La forma ideal de cognición es la automaticidad, por lo que cuanto antes podamos deshacernos de la conciencia, mejor. Entonces, ¿cómo pasa a ser inconsciente la cognición cortical?

Veamos el ejemplo más sencillo posible. ¿Qué creen que pasaría si se proyectara una línea vertical delante de los ojos para eludir los constantes movimientos oculares que normalmente garantizan que todo lo que miramos se actualice tres veces por segundo? Esta actualización se produce en la periferia de la jerarquía predictiva, donde se hallan los mayores grados de libertad. El ritmo frenético de actualización de los modelos que se requiere en circunstancias tan impredecibles hace que la palabra predicción carezca prácticamente de sentido. Por eso se realiza tanto trabajo cognitivo en la periferia sensoriomotora.

¿Qué pasaría si fuera posible inmovilizar un estímulo visual de tal modo que lo viéramos solo y exclusivamente de forma absolutamente monótona? La inquietante respuesta es que desaparece de la conciencia. Aunque el estímulo sigue ahí, desaparece de la conciencia visual en unos segundos. Así lo demostraron en la década de 1950 Lorrin Riggs en Estados Unidos y Robert William Ditchburn en Inglaterra. Desde entonces, se han observado efectos similares en otras modalidades sensoriales.[353] A estas alturas, la razón por la que el estímulo se desvanece de la conciencia ya debería resultarles obvia: se vuelve cien por cien predecible y, por lo tanto, no aporta ninguna información. La predicción alcanza la precisión total y el valor de error se reduce a cero. Esto, como he dicho, es el ideal homeostático del cerebro.

Es difícil estabilizar los estímulos en relación con los movimientos de los ojos y la cabeza, por lo que no se puede hacer el experimento de Riggs-Ditchburn en casa. No obstante, se puede buscar en Google la ilusión del «perseguidor del lila», que muestra otro tipo de desvanecimiento visual.[354] Con la imagen delante, se mantiene la cara a unos veinte centímetros de la pantalla del ordenador, concentrándose en la cruz situada en el centro de los círculos que giran. A continuación, algo empieza a ocurrir en los círculos. Desaparecen porque la corteza visual infiere que son ruido y les asigna menos precisión en relación con el fondo gris, lo que hace que se desvanezcan.

Esta imagen produce otras dos ilusiones visuales. Además del desvanecimiento, vemos cómo un círculo lila se convierte en un círculo verde que no estaba ahí —ni está ni es realmente verde—,

girando alrededor del punto central. Tenemos también una explicación de cómo funcionan estos mecanismos. Básicamente tienen que ver con la ponderación de precisión. Todos estos fenómenos ponen de manifiesto que lo que vemos no lo recibimos del exterior, sino que lo genera nuestro cerebro.[355]

Enlacemos esto con la sección del capítulo 5 sobre la actualización de las preconcepciones motoras innatas: reflejos e instintos. Estas predicciones innatas tienen su utilidad, pero no pueden abarcar toda la complejidad del mundo, por lo que es necesario complementarlas aprendiendo de la experiencia. Este aprendizaje requiere conciencia, ya que poco a poco mejoramos la confianza en nuestras predicciones recién adquiridas. El objetivo de todo aprendizaje es automatizar también estas predicciones adquiridas, hacer que se comporten como reflejos e instintos. Aspiramos a forjar nuevas predicciones que sean al menos tan fiables y generalizables como las antiguas. Cuando lo conseguimos, las predicciones adquiridas se automatizan gracias a la consolidación de los sistemas. La consolidación, en este sentido, es lo contrario de la reconsolidación, que deshace las huellas consolidadas disolviendo literalmente las proteínas que las «cablearon».

Esto ocurre hasta el nivel de los sistemas de memoria no declarativa. El objetivo del aprendizaje por experiencia es desplazar el mayor número posible de recuerdos a largo plazo del estado declarativo al no declarativo, ya que declarativo significa «capaz de volver a la conciencia». Así pues, cuando he dicho que la zombificación es el ideal de la cognición, me refería a que una consolidación cada vez más profunda es el ideal del aprendizaje. La memoria no declarativa es la más fiable. Supone la menor cantidad de trabajo. Minimiza la complejidad y es la más generalizable (véase el capítulo 8). En consecuencia, es la más rápida de ejecutar, implica la menor incertidumbre y, por tanto, el menor retardo.

Por supuesto, la cosa es más compleja. En primer lugar, no quiero dar la impresión de que toda consolidación funciona pasando de los sistemas de memoria declarativa a los sistemas de memoria no declarativa. Muchas predicciones a largo plazo se consolidan directamente en la memoria no declarativa, y casi todos los tipos de aprendizaje se producen simultáneamente en ambos sistemas. En segundo lugar, hay múltiples tipos de memoria no declarativa, y no todos funcionan de la misma forma. Por ejemplo, el aprendizaje «procedimental» tiene lugar mediante la repetición bruta, de ahí que digamos que habilidades y hábitos como montar en bicicleta son «difíciles de aprender y difíciles de olvidar». Sin embargo, algunas variedades de respuesta emocional no declarativa, que son igualmente

difíciles de olvidar, se adquieren mediante un aprendizaje de exposición única: el condicionamiento del miedo, por ejemplo. (Otros tipos de aprendizaje emocional son más lentos, como el vínculo del apego, que requiere unos seis meses). La memoria no declarativa se caracteriza por ser difícil de olvidar, pero la consolidación implica procesos muy diferentes en los distintos sistemas de memoria. Por último, los recuerdos no declarativos solo son «inconscientes» en el sentido cognitivo. Cuando se desencadena una respuesta emocional adquirida, sentimos algo, pero no sabemos a qué se debe, es decir, de dónde viene lo que sentimos (véase el capítulo 3).

Lo más importante de la memoria no declarativa es justamente que no es declarativa. Genera respuestas procedimentales, mientras que la memoria declarativa genera imágenes experimentadas.[356] Esta distinción coincide con una distinción anatómica: los recuerdos declarativos son corticales, mientras que los no declarativos son subcorticales.[357] Las huellas mnémicas subcorticales no pueden recuperarse en forma de imágenes por la razón de que no son mapeos corticales de los órganos finales sensoriomotores. Implican estereotipos más simples, como el que he descrito en este libro en relación con los comportamientos aprendidos de los niños hidranencefálicos y los animales decorticados.[358] Estas cosas no se pueden traer a la mente, no son «pensables».

Los sistemas de memoria cortical, por el contrario, siempre están listos para revivir los escenarios predictivos que representan; literalmente, para volver a experimentarlos. En otras palabras, la memoria declarativa restituye fácilmente las huellas a largo plazo al estado a corto plazo de la memoria de trabajo consciente. Lo hace no solo para actualizar las predicciones corticales, sino también para guiar la acción en condiciones de incertidumbre.

Las huellas mnémicas subcorticales son más fiables que las corticales —sus valores de alta precisión tienen menos probabilidades de cambiar— porque están optimizadas para la simplicidad más que para la precisión. Esto las hace más generalizables. Sin embargo, todo tiene un precio: los modelos menos complejos son menos precisos cuando varía el contexto.[359] En cambio, la relativa complejidad de las predicciones corticales coincide con una mayor plasticidad. Es decir, la corteza se especializa en contextos: restablece la precisión del modelo en situaciones impredecibles.[360] Una cosa compensa a la otra. A mayor potencial de experiencia consciente, menor automatismo, lo que significa más plasticidad pero también más trabajo cognitivo. Eso cuesta energía y genera sensaciones y sentimientos, así que el cerebro lo hace en la cantidad mínima

necesaria. Incluso hasta el punto de hacer desvanecer un estímulo que tenemos ante los ojos.

Sin embargo, muchas de las cosas que nos pasan por la cabeza parecen difíciles de conciliar con este ideal de eficiencia informativa y termodinámica. Junto con los sentimientos y las percepciones, lo que más llama la atención de nuestra conciencia son los pensamientos. Es evidente que su origen es cortical, pero ¿qué son? ¿Y por qué a menudo parecen tan ociosos?

La teoría de la cognición que he estado esbozando aquí gira en torno a la capacidad de nuestros sistemas de memoria para generar un mundo virtual.[361] Cada nivel de la jerarquía predictiva (incluido cada nivel de procesamiento cortical) es capaz de generar una explicación de los datos que espera recibir del siguiente nivel. Esto significa que la percepción no difiere esencialmente de la imaginación: desde el punto de vista subjetivo, hay poca diferencia entre el mundo que se experimenta en sueños y el que hay al otro lado de la ventana.[362] El cerebro puede conjurar realidades fantasmales a la carta. Es de suponer que lo esté haciendo ahora mismo, mientras leemos estas palabras y pensamos en lo que significan. Incluso cuando dejamos vagar la mente…

Esa existencia de una mente vagabunda podría parecer contradictoria con la teoría de la energía libre de la conciencia. He dicho que adquirimos conciencia solo de las señales de error intensas y priorizadas (las salientes), a las que debemos responder si queremos mantener nuestros parámetros biológicos dentro de unos límites viables. Sin embargo, nuestros pensamientos a menudo parecen aleatorios e intrascendentes. En algunos casos, incluso puede que nuestro propio monólogo interno nos resulte intrusivo o nos distraiga, que no ayude nada en las circunstancias del momento. ¿Cómo puede ser que eso minimice la energía libre?

El objetivo del vagabundeo mental, por extraño que parezca, es mejorar la eficiencia de nuestro modelo generativo. Como ordena el principio de la energía libre, un modelo solo es eficiente si utiliza el mínimo de recursos necesarios para realizar el trabajo de autoorganización. Eso se reduce a encontrar el modelo más simple que prediga con éxito muestras sensoriales del mundo. (No olvidemos la navaja de Ockham).

El modelo más sencillo no surge de la acción voluntaria de forma

natural, porque una acción voluntaria es un proceso desordenado. Por tanto, para aumentar la simplicidad se recortan las conexiones sinápticas redundantes que se formaron mientras aprendíamos de la experiencia. La intención es evitar los «sobreajustes» de unos modelos de datos que son ruidosos, pues preservan innecesariamente correlaciones excéntricas y débiles. Las tijeras de podar son los ya conocidos mecanismos de consolidación y reconsolidación de la memoria: al activar los recuerdos, podemos reforzarlos, alterarlos e incluso borrarlos.

El vagabundeo mental es uno de los medios para conseguirlo. Se trata de la actividad espontánea del prosencéfalo (también conocida como «estado de reposo» o «modo por defecto») que se produce en ausencia de cualquier estímulo externo específico. Este tipo de actividad casi siempre tiene lugar en segundo plano, a través de una «exploración imaginativa de nuestro propio espacio mental».[363] Existe un gran solapamiento entre esta forma de pensamiento y el sueño,[364] que parece darse en todas las criaturas dotadas de corteza cerebral: cualquier animal con capacidad para generar imágenes de sí mismo actuando en el mundo también puede deambular por infinitos mundos simulados, según le permitan las circunstancias.[365] El deambular, recordarán, está estrechamente ligado a la pulsión de BÚSQUEDA, que no cesa en sus demandas mientras dormimos. Resulta obvio por qué la actividad en modo por defecto es más segura por la noche: porque no tenemos que lidiar con acontecimientos externos.

Todo esto explica el hecho peculiar de que nuestros modelos adquiridos del mundo nunca sean del todo estables, ni siquiera cuando dormimos. Incluso en ausencia de estímulos sensoriales convincentes, la actividad neuronal estructurada no se interrumpe, lo que da lugar a un proceso continuo de exploración y prueba del modelo generativo. Andy Clark sospecha que estas exploraciones podrían dar lugar a nuevas y elegantes respuestas a problemas que han estado ocupando nuestra atención durante la vigilia, respuestas que a menudo son más sencillas —en el sentido de Ockham, es decir, más eficientes— que nuestros mejores intentos anteriores: «¿Podría todo esto ser al menos parte de la solución a profundos y permanentes enigmas sobre los orígenes de las nuevas ideas y la resolución creativa de problemas?».[366]

La conciencia desempeña el mismo papel en este proceso autogenerado que en la percepción y el aprendizaje a través de la experiencia. Lo que tienen en común todos los procesos cognitivos conscientes es que conllevan el necesario trabajo mental de reconsolidación, es decir, la reducción de predicciones consolidadas a

estados de incertidumbre. Por eso los sueños (que son una forma de resolver problemas) son conscientes.[367]

Sin embargo, hay otros tipos de pensamiento, además del vagabundeo mental. Veamos un segundo tipo: la imaginación deliberada. Si algún proceso cognitivo merece el papel de antagonista en la oposición tradicional entre pensamiento y obra, es este. Aquí pensamos en lugar de actuar, inhibiendo nuestros impulsos motores mientras el sistema tantea los problemas en la imaginación. Cuando pensamos de esta manera, ajustamos nuestras precisiones para suprimir los errores de predicción.[368] Después de todo, el objetivo del pensamiento deliberativo es imaginar que se hacen cosas para calibrar de antemano las posibles consecuencias de hacerlas realmente. (En el capítulo 5 utilizo el ejemplo de pegar a mi director).

¿Cómo imaginamos el futuro? En gran medida, de la misma forma en que recordamos el pasado, que resulta ser, más a menudo de lo que queremos admitir, un pasado imaginado. La memoria episódica es un proceso construido en el que las metas y los contextos actuales contribuyen de forma importante a lo que recordamos. Así, el pasado se revive de forma selectiva en relación con las demandas actuales de trabajo predictivo. Una vez más, el señor S., el hombre del «cartucho de memoria perdido», constituye un ejemplo acertado precisamente porque los mecanismos que sustentan el recuerdo normal están especialmente exagerados en su patología. Su memoria episódica es preocupantemente interesada.

Por supuesto, los cartuchos de memoria no existen. En cambio, los sistemas neuronales implicados en los viajes mentales en el tiempo giran en torno al hipocampo, que es crucial para inyectar la cualidad de la perspectiva «personal» en los procesos de memoria cortical normalmente inconscientes.[369] La investigación contemporánea sobre la memoria episódica revela que, de hecho, el hipocampo está tan implicado en imaginar el futuro como en revivir el pasado.[370] Por ello, David Ingvar habla de «recordar el futuro»[371] y Daniel Schacter conceptualiza el hipocampo (junto con el resto de las estructuras cerebrales responsables de la memoria episódica) como el soporte de una «simulación episódica constructiva» del futuro que implica «la recombinación flexible de detalles de acontecimientos pasados en escenarios novedosos». En opinión de Schacter, el sistema de memoria episódica adquiere su valor adaptativo más por su capacidad de imaginar el futuro que por la de recordar el pasado. El encéfalo, concluye, es «un órgano fundamentalmente prospectivo que está diseñado para utilizar información del pasado y del presente con el fin de generar predicciones sobre el futuro».[372] A estas alturas,

todo esto nos debería resultar ya bastante familiar.

El tercer y último tipo de pensamiento que trataré es el pensamiento con palabras. Al parecer, esta capacidad confiere a la cognición humana su característica más singular. Se suele describir la lengua como una herramienta de comunicación. Que lo es. Sin embargo, es ante todo una herramienta de abstracción. Algunos filósofos se refieren a esta «otra» función del lenguaje como supracomunicativa, pero yo prefiero pensar que es precomunicativa; es difícil imaginar cómo podría haber surgido el habla (por oposición a la vocalización) sin abstracción. El lenguaje no se limita a expresar pensamientos que ya tenemos, también formula pensamientos nuevos.

Un experimento de Gary Lupyan y Emily Ward revela esta función del lenguaje. Utilizaron una técnica denominada «supresión de destello continuo» —otro tipo de percepción biestable, similar a la rivalidad binocular— en la que una imagen presentada de continuo a un ojo queda suprimida de la conciencia por efecto de un flujo cambiante de imágenes presentadas al otro ojo.[373] En este experimento, se mostraba a los sujetos la imagen de un objeto familiar, como una silla, una calabaza o un canguro, en un ojo, mientras que el otro ojo veía una serie de garabatos. Los garabatos suprimieron la imagen estable de la conciencia. Sin embargo, inmediatamente antes de mirar los garabatos y el objeto, los sujetos oían una de estas tres cosas: (1) la palabra para el objeto suprimido (por ejemplo, calabaza cuando el objeto era una calabaza); (2) la palabra para un objeto diferente (por ejemplo, canguro cuando el objeto era una calabaza); (3) solo ruido estático. Cuando se les pidió que indicaran si habían visto algo o no, los sujetos tenían muchas más probabilidades de decir que habían visto conscientemente el objeto estable cuando la palabra que habían oído coincidía con dicho objeto que cuando oían una palabra no coincidente o ninguna palabra. De hecho, oír la palabra diferente reducía todavía más las probabilidades de ver el objeto suprimido. La explicación es que «cuando la información asociada a las etiquetas verbales coincide con la actividad entrante (ascendente), el lenguaje proporciona un impulso descendente a la percepción, enviando a la conciencia una imagen que de otro modo sería invisible».[374] En otras palabras, aumenta la ponderación de la precisión de la imagen perceptiva.

Para comprender el poder de este mecanismo, hay que tener en cuenta que (en realidad, y no solo en este experimento) nos imponemos a nosotros mismos ese etiquetado descendente todo el tiempo a través de los procesos de «discurso interior». El impulso que este discurso proporciona es una forma de imprimación no declarativa.

Curiosamente, la imprimación con palabras tiene un efecto mucho más fuerte en la conciencia que la imprimación con imágenes concretas. Al final resulta que una imagen vale algo menos que una palabra, por no hablar ya de mil. Es de suponer que esto se debe a que las abstracciones (que residen a mayor profundidad en la jerarquía predictiva) pueden lograr más que las imágenes en cuestiones como la «perrunidad», por ejemplo: Lupyan descubrió que oír la palabra perro tiene una probabilidad significativamente mayor de superar la supresión de destellos continuos que el mero hecho de oír ladridos.[375] La abstracción tiene mayor alcance. Así, cuando se incita implícitamente a unos sujetos a prestar atención a los «vehículos» frente a los «seres humanos» mientras ven un videoclip, la imprimación verbal altera la sintonización de poblaciones neuronales enteras, haciéndolas más sensibles a la presencia de una clase de objeto frente a la otra.[376] Las palabras tienen el poder de potenciar categorías semánticas enteras. De hecho, muchas de esas categorías no serían concebibles —ni, por lo tanto, perceptibles— sin sus etiquetas verbales. Esto se aplica de forma más evidente al tipo de conceptos abstractos que estamos considerando en este libro. ¿Quién ha visto alguna vez la energía libre? Sin embargo, una vez que podemos pensar «energía libre», podemos ver su funcionamiento en todas partes.

Si eso es lo que pueden hacer una o dos palabras, ¿qué ocurre cuando empezamos a juntar cientos de ellas? Tomemos, por ejemplo, el potencial de las narrativas personales, las historias que abstraemos y nos contamos a nosotros mismos sobre el flujo y el significado de nuestra vida. «Estas narrativas —escribe Andy Clark— funcionan como elementos de alto nivel en los modelos que estructuran nuestras propias autopredicciones y, por tanto, informan de nuestras acciones y elecciones futuras».[377]

Por supuesto, estas narrativas suelen construirse conjuntamente con otras personas y a lo largo de toda la vida, empezando por el binomio madre-bebé. Así se introduce la función comunicativa de la lengua. La manipulación artificial de la precisión no tiene por qué provenir solo de nuestros propios modelos generativos; también puede provenir de los modelos de otras personas, si tienen capacidades de abstracción similares. Clark coloca esta segunda función del lenguaje bajo el epígrafe «predicción recíproca continua», mientras que Andreas Roepstorff y Chris Frith hablan de «compartir guiones» y de «control superior de la acción».[378] En resumen, el etiquetado abstracto que controla la precisión de una persona puede transmitirse directamente a la de otra, evitando la laboriosa tarea del aprendizaje ascendente.

Roepstorff y Frith lo explican comparando los efectos de dar

instrucciones verbales a otros seres humanos para ayudarlos a realizar una tarea con el arduo proceso de entrenamiento necesario para que un mono adquiera los conocimientos suficientes para hacer lo mismo. En el ejemplo que utilizan Roepstorff y Frith (un juego de clasificación de cartas), los seres humanos adquieren el comportamiento previsto tras unos minutos de instrucción verbal, mientras que los monos tardan un año entero de condicionamiento operativo en captarlo. Sabemos por las pruebas de IRMf que los humanos y los monos de este experimento recurrieron a regiones cerebrales equivalentes (el mismo conjunto cognitivo) para realizar la tarea real. Solo difería el método de adquisición.

El lenguaje implica algo más. Abre la puerta a toda una serie de técnicas de mejora de la precisión, como las observaciones, las teorías y las ecuaciones que presentamos en este libro. Gracias a las palabras (y a otros símbolos, como los matemáticos), los modelos adquiridos a lo largo de una vida individual pueden convertirse en objetos estables de análisis y mejora sistemática por parte de otros, no solo nuestros contemporáneos, sino a lo largo de generaciones. Con sus maravillosas gradaciones de características genéricas y específicas, el lenguaje nos permite proyectar en la conciencia algo de la estructura de la propia jerarquía predictiva. Las especies no simbólicas no pueden acceder a estas poderosas ayudas a la cognición. Es muy difícil imaginar el conjunto de la ciencia, la tecnología y la cultura sin la lengua.

Lo que sospecho yo es que el lenguaje evolucionó principalmente a partir de la pulsión de JUEGO. En el capítulo 5 explicaba cómo la pulsión de JUEGO da lugar a la formación de normas sociales. Estas normas regulan el comportamiento del grupo y así nos protegen de las necesidades potencialmente excesivas de cada individuo. La regla 60-40 es una regla social innata. Exige reciprocidad y mutualidad, y, por tanto, facilita el desarrollo de la empatía, es decir, la capacidad de leer otras mentes. Es fácil ver cómo la forja de reglas sociales adquiridas fomenta formas complejas de comunicación para expresarlas y como todo ello contribuye a su vez a la aparición del pensamiento simbólico. La presión para desarrollar normas artificiales aumentó exponencialmente cuando los humanos abandonaron los antiguos estilos de vida de cazadores-recolectores típicos de todos los primates y empezaron a vivir en asentamientos permanentes, basados en la agricultura y la ganadería. No teníamos ninguna preparación evolutiva para este desarrollo, que se produjo hace apenas doce mil años, es decir, ninguna salvo la pulsión de JUEGO, que es fundamental en la formación de jerarquías sociales.

Por lo tanto, es de gran interés observar que la corteza cerebral

contribuye más al JUEGO que a cualquier otra emoción básica.[379] Su cualidad de «como si» es imposible de concebir sin mecanismos corticales del tipo que he descrito en este capítulo. La función de JUEGO bien podría ser un precursor biológico del pensamiento en general (es decir, de toda la acción virtual frente a la real) y de toda la vida cultural.

Los lectores perspicaces notarán que lo que acabo de decir también sugiere algo sobre el papel de la cura mediante la palabra en Freud. ¿Qué es la terapia conversacional sino una intervención específica en la narrativa personal? En mi opinión, la psicoterapia —una forma de «predicción recíproca continua»— es también una forma de JUEGO. En fin, estoy entrando en temas que merecerían libros propios. Ahora voy a terminar esta reflexiones sobre la corteza cerebral considerando la diferencia esencial entre la conciencia cortical y la del tronco encefálico.

A Panksepp le interesaba mucho nuestra tendencia a asociar ciertos colores con determinadas sensaciones.[380] Por ejemplo, describir los colores del espectro rojo como cálidos y los del espectro azul como fríos es algo convencional. ¿Se trata de un mecanismo arbitrario? ¿Podrían invertirse los colores? La calidez y la frialdad exteroceptivas no son, por supuesto, cualidades visuales, como lo son el rojo y el azul. Son características de la sensación somática, pero también están estrechamente relacionadas con lo que Panksepp denominó «afectos sensoriales», como el asco, el dolor y la sorpresa.[381] En varios sentidos, los aspectos hedónicos del calor y el frío encarnan un valor biológico: la proximidad física de los demás es cálida, el sexo es cálido, el fuego es cálido, *etc.* Es posible que los espectros de color asociados al fuego y al hielo hayan adquirido sus significados afectivos de este modo, al igual que los colores de la fruta madura frente a la inmadura. Panksepp especuló con la posibilidad de que estas asociaciones fueran un vestigio de los orígenes evolutivos de los qualia perceptivos, un recuerdo de la época en que nuestras únicas percepciones conscientes eran los afectos sensoriales.

Sin embargo, hay una gran diferencia entre la sensación de los afectos y la valoración asociativa de objetos percibidos externamente, como la fruta madura o las hogueras calientes. La diferencia es la siguiente: los afectos son algo innato, pero la valoración de las percepciones externas es algo adquirido. Las cualidades visuales y otras cualidades perceptivas de las percepciones se «asocian» a los afectos en el sentido empírico.[382] Otorgamos valores al mundo. Aunque algunas de estas asociaciones se ajustan a esquemas regulares que existen en la naturaleza (entre el rojo y el fuego y la madurez de un fruto, por un

lado, y ciertas sensaciones placenteras, por otro), incluso esas asociaciones son adquiridas y, por tanto, están sujetas a los caprichos de la experiencia individual.

En resumen, estas asociaciones son contextuales. Por ejemplo, el fuego no siempre es bueno, por eso los colores del espectro cálido, como el rojo, a veces denotan cosas buenas, como el sexo, y a veces denotan cosas malas, como el peligro. Por eso no solo hablamos de sexo «caliente», sino también de «agresión en caliente». Este principio se aplica en mayor grado a medida que se individualizan las asociaciones que constituyen la iconografía subjetiva de cada persona. Es la lógica de Lisa Feldman Barrett.

La opinión de Panksepp era que los valores afectivos asociados a algunas cualidades perceptivas son un vestigio de sus orígenes en los afectos sensoriales. No obstante, no debemos perder de vista que la conexión innata se ha roto cuando vinculamos cualidades perceptivas como el enrojecimiento o el calor con sentimientos eróticos, y el azul o la frialdad con sentimientos tristes. Por eso es tan posible condicionar las asociaciones color-afecto.

Sin duda, la presencia regular en una población de estas asociaciones, que hasta cierto punto están condicionadas por la naturaleza, tiene que constituir la base de formas artísticas como la pintura, la música y la danza. Apreciamos sus cualidades visuales, auditivas y somatosensoriales (al menos en parte) a través de nuestros afectos sensoriales y las connotaciones que evocan en nosotros. Siguiendo mi definición de la conciencia como incertidumbre sentida, también es interesante observar el papel de la sorpresa en la experiencia estética. Lo mismo ocurre con el humor: una obra de arte aburrida es como un chiste cuyo final es demasiado evidente. En cualquier caso, estos temas son demasiado amplios para abordarlos bien aquí.[383]

Podemos concluir que los qualia perceptivos son diferentes de los qualia afectivos en que no tienen una valencia hedónica inherente; adquieren su valencia en relación con el afecto. Las ondas luminosas y sonoras que inciden continuamente en la corteza cerebral no siempre se experimentan de forma consciente,[384] pero cuando lo hacen se experimentan como un contexto (véanse pp. 170-172).

En última instancia, la percepción consiste en el cálculo de estadísticas inferenciales relativas a las distribuciones de probabilidad de los trenes de impulsos nerviosos y en comparaciones entre dichas probabilidades, todo ello en una jerarquía anidada de homeostatos de optimización de la precisión.[385] La jerarquía genera una

representación gráfica mental de «respuestas suscitadas por el equipo»; fenómenos en el sentido de John Wheeler. Estos fenómenos constituyen la realidad virtual desplegada ante (y por) cada uno de nosotros. Los estados de reposo de los miles de millones de pequeños homeostatos —todos ellos incrustados unos dentro de otros desde la superficie hasta las profundidades— representan nuestra confianza en un contexto previsible. El contexto es lo que esperamos que ocurra más allá de nuestra manta de Markov cuando intentamos resolver una necesidad priorizada. Dar prioridad a una necesidad desencadena un trabajo cognitivo arraigado en una creencia concomitante: el resultado previsible. Lo que se percibe conscientemente, según el mecanismo que ahora conocemos, no es este contexto previsible, sino el despliegue anidado —en un esquema jerárquico profundo— de desviaciones priorizadas con respecto a nuestras expectativas.

En la periferia de la jerarquía, las fluctuaciones son constantes, es casi imposible predecir el presente en todos sus detalles. Nuestra confianza variable en los aspectos destacados de nuestras expectativas sensoriomotoras se experimenta en forma de colores y tonos y similares, es decir, disposiciones de las variables categóricas exteroceptivas que computa nuestra especie. Esto incluye no solo las percepciones del mundo exterior, sino también las desviaciones de los estados propioceptivos previsibles del avatar que representa nuestro propio cuerpo, ya que la experiencia de nuestro cuerpo no es menos virtual que la experiencia del mundo exterior.[386]

Los errores de precisión de predicción en la percepción y la propiocepción (cuando tienen saliencia) se registran como qualia exteroceptivos. En cambio, la confianza variable en la creencia fenotípica que de entrada generó el proceso sensoriomotor se experimenta en forma de afecto. Sea cual sea el proyecto que estemos llevando a cabo, sentimos a cada momento lo bien o mal que lo estamos haciendo. Como les debe de pasar ahora mismo a quienes leen esto. Al leer estas palabras se produce un cambio en su interior, un flujo de incertidumbre creciente y menguante, que podría frenarles en seco si crece demasiado. Su cualidad de arraigo, el afecto, lo sentimos a lo largo de toda la vida y, en última instancia, regula todo lo que hacemos.

Lo que estoy sugiriendo es que nuestra experiencia cotidiana no consiste en última instancia más que en esto.

[330] Zeki, 1993, p. 236; la cursiva es mía.

[331] Ibid., p. 238.

[332] No quiero decir que se trate de un proceso cognitivo de alto nivel. Es como si la corteza visual de «proyección» rellenara el punto ciego con lo que le rodea. En condiciones naturales, esto se ve favorecido por los frecuentes movimientos de los ojos, que garantizan que casi siempre se tenga una imagen real (con memoria ultracorta) de lo que hay en el punto ciego. Sobre los diferentes mecanismos implicados, véanse Ramachandran, 1992; Ramachandran y Gregory, 1991; y Ramachandran, Gregory y Aiken, 1993. Por cierto, junto al relleno del punto ciego se produce un filtrado de objetos no deseados, como las «moscas volantes» entópticas y la vasculatura retiniana.

[333] El solapamiento entre los campos representados por cada ojo anula los puntos ciegos, lo que significa que es necesario cerrar uno de ellos para observar el relleno ilusorio que acabamos de mencionar.

[334] Por supuesto, existe cierto solapamiento entre ellos, pero el hecho de que los objetos de nuestra atención visual están literalmente bisecados se demuestra fácilmente en los casos de inatención unilateral (izquierda) tras una lesión en el hemisferio derecho.

[335] Esto, por cierto, demuestra que la función «vinculante» de la conciencia proviene de debajo de la corteza, del tronco encefálico unitario y no de las cortezas bicamerales (véase Panksepp y Biven, 2012).

[336] Solms et al., 1988.

[337] Lo describió como «una maravillosa pieza teatral» (Helmholtz, 1867, p. 776).

[338] Ibid., p. 438.

[339] La mayoría de las personas que trabajan con el paradigma de la codificación predictiva no prestan suficiente atención al hecho de que la inferencia perceptiva es un proceso inconsciente. Por ejemplo, no estoy de acuerdo con Hohwy (2013) cuando dice que «de lo que somos conscientes es de la “fantasía” generada por la forma en que las predicciones van atenuando el error de predicción». En mi opinión, no nos hacemos conscientes de nuestras «fantasías» predictivas a menos que choquen con la realidad. Y menos mal… Parece que Clark está de

acuerdo conmigo en este aspecto (Lupyan y Clark, 2015, p. 281; la cursiva es mía): «Aunque la mayoría de las predicciones son inconscientes, a veces uno puede ser consciente de ellas cuando no se cumplen. Por ejemplo, imaginemos beber de un vaso lo que nos parecía zumo de naranja y darnos cuenta al probarlo de que en realidad era leche. La diferencia entre el sabor de esa leche cuando esperamos leche y cuando esperamos zumo de naranja es la expectativa de zumo de naranja llevada a la conciencia (Lupyan, 2015, para debate). Del mismo modo, consideremos la experiencia de una omisión inesperada, como cuando falta una nota musical en una composición conocida. Estas omisiones pueden ser tan llamativas desde el punto de vista perceptivo y tan salientes como el tono más vibrante, un efecto por lo demás desconcertante que se explica claramente cuando consideramos que la construcción de la experiencia perceptiva implica expectativas basadas en algún tipo de modelo de lo que es probable que ocurra».

A pesar de lo que Hohwy (2013) dice, a veces se acerca a mi punto de vista. En un artículo anterior hizo la siguiente afirmación (Hohwy, 2012, p. 11; la cursiva es mía): «Esta firma temporal es coherente con la codificación predictiva, en la medida en que cuando el error de predicción de un estímulo se suprime de forma generalizada y no hay más exploraciones (ya que la inferencia activa está oculta debido a la fijación central durante la atención encubierta), la probabilidad debería empezar a caer. Esto se deduce de la idea de que lo que impulsa la percepción consciente es el proceso real de supresión del error de predicción». Sin embargo, en su libro (2013, p. 201; la cursiva es mía) dice: «La percepción consciente es el resultado de la inferencia perceptiva inconsciente. La actualización bayesiana de nuestras probabilidades anteriores a la luz de nuevas pruebas no es algo consciente, como no lo es la forma en que se predice y luego se atenúa el aporte sensorial. Lo que es consciente es el resultado de la inferencia, la conclusión». Por eso no estoy de acuerdo con Hohwy (2013), si no le he entendido mal. En lo que estamos de acuerdo es en la opinión de que la conciencia la genera el trabajo de minimizar las señales de error precisas, es decir, el problemático desajuste entre la predicción y el error con ponderación de precisión. En cambio, no estoy de acuerdo con él cuando dice que «de lo que somos conscientes es de la “fantasía”». En el mejor de los casos, Hohwy parece pensar que somos conscientes de la «fantasía» que intenta dar por explicado el error de entrada, pero yo creo que somos conscientes del hecho de que no está suprimiendo el error, es decir, somos conscientes del «trabajo en curso» de predicción causado por el desajuste. Eso es lo que hace saliente la realidad. Quizá se trate solo de una cuestión

semántica. Para mí, lo esencial es que la conciencia en la percepción está impulsada por la incertidumbre, no por las mejores suposiciones a las que da lugar la certeza relativa. A mi modo de ver, las suposiciones se hacen conscientes solo cuando son inciertas, y se repliegan cuando se confirman. En otras palabras, solo nos hacemos conscientes de nuestras fantasías cuando la realidad las contradice. La conciencia podría describirse como un proceso de desambiguación.

Soy consciente de que esta cuestión se complica por la diferencia entre inferencia perceptiva e inferencia activa (y entre atención exógena y endógena). Esto explica por qué hay tantos trabajos publicados (los trabajos en los que se basa Hohwy, 2013) que sugieren que percibimos preferentemente (conscientemente) aquello que se ajusta a nuestras expectativas, y también por qué hay otros tantos trabajos que sugieren que percibimos preferentemente (conscientemente) aquello que es más inesperado. Para mí, esta contradicción se resuelve mediante el concepto de priorización de los afectos, que en ambos casos se rige por el equilibrio entre lo que he denominado vagamente «necesidades» y «oportunidades» (véase el capítulo 5). En resumen, percibimos conscientemente aquello que destaca más (saliente) en relación con nuestra necesidad priorizada en ese momento. Por definición, las necesidades priorizadas (afectos) producen las señales de error más precisas.

Mi opinión sobre esta cuestión se desarrolla en la siguiente sección. Sin embargo, en última instancia se trata de una cuestión empírica. Junto con mis alumnos Donne van der Westhuizen y Julianne Blignaut y mi exalumno Joshua Martin, investigo la cuestión utilizando el paradigma estándar de rivalidad binocular, que es un paradigma relativo a la conciencia perceptiva. Véanse también Pezzulo, 2014; Yang, Zald y Blake, 2007; y Stein y Sterzer, 2012.

[340] Publiqué esta idea por primera vez en Solms, 2013, y luego la he desarrollado en diversas direcciones, por ejemplo, Solms, 2015b, 2017b, 2017c, 2018b.

[341] Freud, 1920, p. 25; la cursiva es mía.

[342] Freud asignó la conciencia y la memoria a dos sistemas diferentes de neuronas (ω y ψ, respectivamente), que más tarde se convirtieron en sus sistemas metapsicológicos Cc. y Pcc., pero los interpretó a ambos como sistemas corticales. El pasaje del que se cita la frase «la conciencia surge en remplazo de la huella mnémica» lo deja muy claro, a pesar de que se escribió veinticinco años después del «Proyecto» (véase Freud, 1920, pp. 25 y ss.).

[343] Bargh y Chartrand, 1999, p. 476: «Algunos de los sistemas de orientación automática que hemos descrito son “naturales” y no requieren experiencia para su desarrollo. Se trata de la confraternización de las representaciones perceptivas y conductuales y de la conexión entre los procesos de evaluación automática, por un lado, y el estado de ánimo y el comportamiento, por otro. Otras formas de autorregulación automática se desarrollan a partir de la experiencia repetida y constante; se adaptan a las regularidades de la propia experiencia y toman el relevo de la elección y la orientación conscientes cuando esa elección no se está ejerciendo. Así es como las metas y los motivos pueden llegar a operar de forma no consciente en determinadas situaciones, como los estereotipos pueden llegar a asociarse de forma crónica con las características perceptivas de los grupos sociales y como las evaluaciones pueden llegar a integrarse en la representación perceptiva de la persona, el objeto o el suceso, de modo que se activen de forma inmediata e involuntaria en el transcurso de la percepción».

[344] Técnicamente, los estados neuronales, potenciados por las precisiones neuronales, actualizan los parámetros neuronales.

[345] Esto coincide exactamente con la noción de Freud de «proceso secundario». Freud conceptualizó el proceso secundario como una «vinculación» de la forma primaria de energía motriz, que es «libremente móvil». La vinculación de la energía libre de Friston es el fundamento mecánico de lo que llamamos trabajo mental (efectivo). Esta conclusión es muy significativa: «En mi opinión, esta distinción [entre energía libre y energía ligada] representa el conocimiento más profundo que hemos adquirido hasta ahora sobre la naturaleza de la energía nerviosa, y no veo cómo podemos prescindir de él» (Freud, 1915b, p. 188). No obstante, cabría destacar que Freud consideraba que el proceso secundario también puede funcionar de forma preconsciente. Esto nos lleva a una pregunta empírica: ¿puede el procesamiento cortical desempeñar una función estabilizadora en ausencia de conciencia? Esta pregunta hace pensar en la controversia actual sobre la memoria de trabajo no declarativa (véase Hassin et al., 2009).

[346] No quiero decir que toda reconsolidación implique la conciencia del recuerdo que está siendo actualizado; también puede implicar la modulación afectiva del proceso de actualización que permanece cognitivamente inconsciente (véase más adelante). Por cierto, los mecanismos que acabo de repasar explican el fenómeno psicológico que Freud denominó «resistencia», nuestra especial reticencia a actualizar nuestros modelos predictivos ante pruebas contradictorias.

(¡Por desgracia, es así incluso para los científicos!).

[347] El término «recompensa = predicción» en el diagrama de Schultz puede confundir a algunos lectores. Cuando las consecuencias sensoriales de una acción coinciden con las consecuencias predichas, no ocurre nada, no hay «recompensa». Los sentimientos, tanto los negativos como los positivos, siempre suponen un error (véase fig. 12). Sin embargo, en el lenguaje conductista, recompensa no implica ningún sentimiento. Simplemente, es un refuerzo de la predicción. Es decir, usando mi terminología, se le asigna una mayor precisión. Para mí, eso significa que se percibirá como algo placentero si la predicción pertinente ha sido priorizada por el triángulo decisorio del mesencéfalo.

[348] Misanin, Miller y Lewis, 1968.

[349] Nader, Schafe y LeDoux, 2000. Véase Dudai, 2000, para una perspectiva general accesible. La reconsolidación está estrechamente relacionada con el concepto de Freud de «retranscripción» de los recuerdos. Véase su carta a Fliess de 6 de diciembre de 1896, en la que presagia, como mínimo, el concepto de consolidación de sistemas: «Como usted sabe, estoy trabajando partiendo de la base de que nuestro mecanismo psíquico ha surgido mediante un proceso de estratificación: el material presente en forma de huellas mnémicas es sometido de vez en cuando a una reorganización acorde con las nuevas circunstancias, a una retranscripción. Así pues, lo esencialmente nuevo de mi teoría es la tesis de que la memoria está presente no una sino varias veces, que se establece en diversos tipos de indicios. Hace tiempo postulé un tipo similar de reordenación (Aphasia [Freud, 1891]) para los caminos que llevan allí desde la periferia. No puedo decir cuántos de estos registros existen: al menos tres, quizá más. Esto se puede ver en el esquema siguiente [donde la conciencia aparece como el "registro" final de la huella; véase fig. 8], que parte de la base de que los diferentes registros también están separados (no necesariamente desde el punto de vista tópico) según sus portadores neuronales. Esta suposición puede no ser necesaria, pero es la más sencilla y podemos darla por buena provisionalmente». Freud conceptualizó la «represión» como un fracaso de la retranscripción. Véase en Solms, 2017c una actualización neuropsicoanalítica.

[350] De ahí que tanto la potenciación a largo plazo como la depresión a largo plazo estén moduladas por el sistema activador reticular (véase Bienenstock, Cooper y Munro, 1982). De ahí también la capacidad de la terapia electroconvulsiva y de los ataques

epilépticos, ambos actuando a través del sistema activador reticular, para interferir en la consolidación de la memoria.

[351] Lo que Hebb (1949, p. 62) dijo en realidad fue: «Supongamos que la persistencia o repetición de una actividad reverberante (o "huella") tiende a inducir cambios celulares duraderos que aumentan su estabilidad. [...] Cuando un axón de la célula A está lo suficientemente cerca como para activar una célula B y participa repetida o persistentemente en su activación, se produce un proceso de crecimiento o cambio metabólico en una o ambas células, de tal manera que la eficacia de A, como una de las células que activan B, aumenta».

[352] Estoy planeando, junto con Cristina Alberini, una serie de experimentos sobre el papel de la sustancia gris periacueductal y del sistema reticular activador en el aprendizaje consciente frente al inconsciente. Estos experimentos aclararán las funciones de los distintos núcleos de «excitación» en el tronco encefálico superior que modulan la cognición consciente en el prosencéfalo. El paradigma de la reconsolidación promete revelar algunos de los mecanismos intracelulares elementales de la conciencia perceptiva en relación con el aprendizaje. Por ejemplo, tanto la consolidación como la reconsolidación pueden verse alteradas por la inhibición de la síntesis de las proteínas, y ambas requieren el factor de transcripción genética CREB. Sin embargo, investigaciones recientes sugieren que en la amígdala es necesario el BDNF para la consolidación, pero no para la reconsolidación, y que el factor de transcripción y gen inmediato temprano Zif268 es necesario para la reconsolidación, pero no para la consolidación. En el hipocampo se encontró una doble disociación similar entre Zif268 para la reconsolidación y BDNF para la consolidación. Véanse Debiec et al., 2006; y Lee, Everitt y Thomas, 2004.

[353] Riggs y Ratliff, 1951; Ditchburn y Ginsborg, 1952. Resulta instructivo recordar las observaciones de Helmholtz sobre la atención (1867, p. 770): «El estado natural no forzado de nuestra atención es vagar hacia cosas siempre nuevas, de modo que cuando se agota el interés de un objeto, cuando no podemos percibir nada nuevo, entonces la atención, contra nuestra voluntad, se dirige a otra cosa [...]. Si queremos que la atención se fije en un objeto, tenemos que seguir encontrando algo nuevo en él, sobre todo si otras sensaciones intensas tratan de desvincularnos de él». Esto debe relacionarse con lo que he dicho antes sobre la BÚSQUEDA por defecto.

[354] https://es.wikipedia.org/wiki/Perseguidor_del_lila.

[355] Si alguien pretende que estos efectos no implican al cerebro (es decir, que solo se producen en niveles más periféricos del sistema nervioso), habría que recomendarle leer a Hsieh y Tse, 2006. Véase también Conant y Ashby, 1974.

[356] Oberauer et al., 2013.

[357] Excluyo aquí la «imprimación» y el aprendizaje perceptivo. Estos dos procesos sí conllevan imágenes e implican a la corteza cerebral. La mayoría de las demás cosas que digo sobre la memoria no declarativa no se aplican a la imprimación ni al aprendizaje perceptivo, que pertenecen a una categoría propia; son el «andamiaje» de la memoria declarativa. Véase lo que digo más adelante sobre la función de imprimación de las palabras, por ejemplo.

[358] Los ganglios basales están preservados en algunos niños hidranencefálicos (pero no en todos) y en los animales decorticados.

[359] Técnicamente, la complejidad es la entropía relativa entre las creencias posteriores y anteriores o las distribuciones de probabilidad sobre estados externos. Esta definición de complejidad se desprende del hecho de que la evidencia del modelo es la diferencia entre precisión y complejidad (véase el capítulo 8). A medida que se incrementa activamente la evidencia del modelo minimizando la energía libre, aumenta la precisión de las predicciones, con el aumento concomitante de la complejidad. En otras palabras, el aumento de la complejidad del modelo siempre se deriva de la capacidad de hacer predicciones más precisas, como suele ocurrir en los sistemas de memoria cortical.

[360] Se puede comparar con Hohwy, 2013, p. 202: «La idea sería que la inferencia perceptiva se mueve en un espacio determinado tanto por la exactitud del error de predicción como por la precisión del error de predicción. Esto se puede representar de forma simplificada si concebimos la exactitud del error de predicción como algo que aumenta con amplitud inversa a la del propio error de predicción, y la precisión del error de predicción, como algo que aumenta con amplitud inversa a la de las fluctuaciones aleatorias en torno a la incertidumbre sobre las predicciones [...]. Esto apunta a una explicación unificada de la relación entre percepción consciente y atención. Se consideran, una respecto a la otra, como una inferencia estadística de primer orden y de segundo orden». Véase también Hohwy, 2012, ya comentado brevemente.

[361] Esta sección se basa en gran medida en Clark, 2015, por lo que

le damos las gracias.

[362] La diferencia se basa en el hecho de que los sueños están casi desprovistos de señales de error exteroceptivas, debido a los drásticos cambios en la ponderación de la precisión que se producen con el inicio del sueño (Hobson y Friston, 2012, 2014). Clark lo describe como «aislamiento de la inducción», pero véase la nota 35.

[363] Clark, 2015, p. 273.

[364] Sin embargo, hay que tener en cuenta que el cerebro dormido no está «aislado de la inducción», como dice Clark, 2015, p. 107. La principal «señal sensorial motriz» del cerebro es siempre endógena. Sencillamente, no es posible aislarse de la inducción de esta señal, es decir, de nuestras necesidades biológicas. Si los neurocientíficos seguimos pasando por alto este hecho fundamental, nunca comprenderemos la vida mental y su lugar en la naturaleza.

[365] Domhoff, 2017.

[366] Clark, 2015, p. 274.

[367] Véase Solms, 2020a, para profundizar en este punto.

[368] Es algo que ocurre también en el «estado de reposo» por defecto.

[369] Es fácil olvidar que el hipocampo forma parte del sistema límbico, el cerebro emocional.

[370] Okuda et al., 2003; Szpunar et al., 2007; Szpunar, 2010; Addis et al., 2007.

[371] Ingvar, 1985.

[372] Schacter, Addis y Buckner, 2007, p. 660; la cursiva es mía.

[373] El hecho de que un flujo cambiante (impredecible) suprima una imagen constante (más predecible) es, por supuesto, interesante en sí mismo.

[374] Lupyan y Ward, 2013, p. 14196; la cursiva es mía. Desde el punto de vista de las cuestiones técnicas relativas a la percepción consciente comentadas anteriormente, se trata, por supuesto, de un caso de atención endógena. A este respecto, véase también la definición técnica de «saliencia» que figura en la p. 245. Las palabras

multiplican artificialmente la saliencia. Sin embargo, la imprimación verbal queda fácilmente anulada por una sorpresa exógena intensa (es decir, precisa). Dicho de otro modo, la priorización ascendente de las necesidades en el nivel del triángulo decisorio del mesencéfalo supera invariablemente a los procesos descendentes del prosencéfalo.

[375] Lupyan y Thompson-Schill, 2012.

[376] Véase de nuevo el experimento de ceguera por falta de atención citado en el capítulo 9, nota 8.

[377] Clark, 2015, p. 286.

[378] Roepstorff y Frith, 2004.

[379] Zhou et al., 2017.

[380] Panksepp y Biven, 2012, p. 396: «Inicialmente, la experiencia de la vista y el oído conscientes era en gran medida afectiva (Panksepp, 1998). La inmediatez con la que los estímulos visuales y auditivos repentinos pueden sobresaltarnos y asustarnos, especialmente cuando dichos estímulos se originan muy cerca de nuestro cuerpo, sugiere una profunda integración primitiva de estos sistemas sensoriales con algunos de nuestros mecanismos afectivos de supervivencia más esenciales. También hay que tener en cuenta que somos propensos a asociar determinados colores con sentimientos».

[381] También hay que tener en cuenta el concepto cada vez más acreditado de «toque afectivo».

[382] No estoy diciendo que la conciencia perceptiva en su conjunto no tenga poder causal. Adquiere su poder precisamente porque contextualiza el afecto.

[383] Véase Hurley, Dennett y Adams, 2011, un libro sobre el humor sobre el que Dennett me llamó la atención recientemente porque llega a conclusiones sobre la arquitectura funcional de la mente que son notablemente similares a las mías.

[384] La corteza cerebral mide y compara continuamente las longitudes de onda y las intensidades auditivas y visuales, y las clasifica de forma tanto consciente como inconsciente.

[385] Cf. Clark, 2015, p. 207: el paradigma de la energía libre sugiere «no que experimentamos nuestras propias señales de error de predicción (o sus precisiones asociadas) como tales, sino que esas

señales actúan dentro de nosotros para captar los flujos aptos de predicciones que revelan un mundo de objetos y causas distales».

[386] Si el uso que hago de la palabra avatar parece alarmante, hay que recordar que todo lo que percibimos es virtual, incluida la imagen que tenemos de nuestro propio cuerpo. Consideremos lo que escribí una vez sobre la ilusión del «intercambio de cuerpos» (Solms, 2013, p. 15): «El sujeto de la conciencia se identifica con su cuerpo externo (objeto-presentación) del mismo modo que un niño se proyecta en la figura animada que controla en un juego de ordenador. La representación es rápidamente invadida por un sentido del yo, aunque no sea realmente el yo. He aquí un sorprendente experimento que ilustra muy claramente la relación contraintuitiva que existe en realidad entre el yo subjetivo y su cuerpo externo. Petkova y Ehrsson (2008) relatan una serie de experimentos de “intercambio de cuerpos” en los que cámaras montadas sobre los ojos de otras personas o maniquíes transmitían imágenes desde ese punto de vista a gafas de videovigilancia montadas sobre los ojos de los sujetos experimentales, creando rápidamente en dichos sujetos la ilusión de que el cuerpo de la otra persona o el maniquí era su propio cuerpo. Esta ilusión era tan convincente que persistía incluso cuando los sujetos proyectados estrechaban la mano a su propio cuerpo. La existencia de la ilusión también se demostró objetivamente por el hecho de que cuando el otro cuerpo (cuerpo ilusorio) y el cuerpo propio (cuerpo real) eran amenazados con un cuchillo, la respuesta de miedo (la “reacción visceral” del cuerpo interno, medida por la frecuencia cardíaca y la respuesta galvánica de la piel) era mayor para el cuerpo ilusorio […]. Esto nos recuerda que la corteza cerebral no es más que memoria RAM». Sobre este último punto, véase Ellis y Solms, 2018.

11

El problema difícil

El físico Paul Davies escribe:

Entre las muchas desconcertantes propiedades de la vida, el fenómeno de la conciencia se caracteriza por ser uno de los más sorprendentes. Es posible que su origen sea el problema más difícil al que hoy se enfrenta la ciencia, el único que sigue siendo casi impenetrable incluso después de dos mil quinientos años de debatir al respecto […]. La conciencia es el principal problema de la ciencia, y hasta podría decirse que de la existencia.[387]

Podría citar aquí muchas afirmaciones similares de otros científicos. El «problema difícil» (como decidieron llamarlo reverencialmente) pregunta por qué y cómo nosotros —«nuestras alegrías y nuestras penas, nuestros recuerdos y nuestras ambiciones, nuestro sentido de la identidad personal y del libre albedrío»,[388] en suma, nuestra experiencia de la existencia— podríamos surgir de los procesos psicológicos que tienen lugar en las células del cerebro, unas células que, en lo fundamental, no se diferencian de las que constituyen otros órganos corporales. Entonces ¿cómo «nos» traen al ser?

No puede decirse que se trata de una pregunta nueva; de hecho, es probable que estemos ante el más antiguo y sentido de todos los misterios humanos. En el pasado se formulaba así: «¿Cómo llega mi alma a residir en mi cuerpo?», pero no adquirió su forma actual hasta 1995, cuando la formuló el filósofo David Chalmers. Permítanme que cite aquí su célebre formulación:

Es innegable que algunos organismos son sujetos de la experiencia, pero lo que desconcierta es la manera en que esos sistemas son sujetos de la experiencia. ¿Por qué cuando nuestros sistemas cognitivos toman parte en el procesamiento de información visual y auditiva tenemos experiencia visual y auditiva: la calidad de azul profundo, la sensación de oír un do central? ¿Cómo podemos explicar por qué hay algo que

es como contemplar una imagen mental o experimentar una emoción? Está ampliamente aceptado que la experiencia surge de una base física, pero carecemos de una buena explicación de por qué y cómo surge. ¿Por qué el procesamiento físico debería dar lugar a una rica vida interior? Desde un punto de vista objetivo, no parece razonable que debiera hacerlo, y sin embargo lo hace.[389]

La formulación de Chalmers está en deuda con un artículo anterior del filósofo Thomas Nagel titulado «¿Cómo es ser un murciélago?» (1974), en el que Nagel hizo hincapié en el carácter de «algo que es como» propio de la experiencia subjetiva y señaló que «un organismo tiene estados de conciencia mental si y solo si hay algo que es como ser ese organismo, algo que es como para el organismo». Y añadió: «Si reconocemos que una teoría física de la mente debe explicar el carácter subjetivo de la experiencia, hay que admitir que ninguna idea hoy día disponible nos ofrece una pista sobre el modo en que podría hacerse». Para concluir: «No parece probable que pueda contemplarse la posibilidad de que exista una teoría física de la mente hasta que se reflexione más sobre el problema general de lo subjetivo y lo objetivo».

Lo que dos décadas después movió a Chalmers a replantear el «problema general» tal como lo hizo fue el hecho de que los neurocientíficos habían empezado a abordar el tema de la conciencia con métodos experimentales. Gracias a los progresos tecnológicos que he descrito en el capítulo 1, creían que ya podían responder a preguntas como esta: «¿Cómo las células del cerebro convierten procesos fisiológicos en experiencias?». Uno de los primeros científicos en abordar el problema con métodos experimentales fue el biólogo molecular sir Francis Crick, codescubridor de la estructura del ADN. Y lo hizo en un libro titulado La búsqueda científica del alma. Una revolucionaria hipótesis para el siglo XXI, publicado apenas un año antes de la trascendental declaración de incredulidad de Chalmers. Crick escribió:

La hipótesis revolucionaria sería que nosotros, nuestras alegrías y nuestras penas, nuestros recuerdos y nuestras ambiciones, nuestro sentido de la identidad personal y del libre albedrío, solo son en realidad el comportamiento de un gran número de células nerviosas y las moléculas a ellas asociadas.[390]

Chalmers no rechazó de plano la afirmación de que la conciencia tiene una base física. Como acabamos de señalar, fue él quien dijo que «está ampliamente aceptado que la experiencia surge de una base física». Chalmers se limitó a afirmar que «carecemos de una buena explicación de por qué y cómo surge». Estamos ante una repetición de lo que había afirmado Nagel, a saber, que «ninguna idea hoy día disponible nos ofrece una pista sobre el modo en que podría hacerse». Crick había afirmado que se podía encontrar una buena explicación, pues la tecnología ahora disponible podía identificar con facilidad lo que él llamó «correlato neuronal de la conciencia». Asimismo, pensaba que si aislábamos las partes anatómicas del cerebro necesarias para la conciencia y las funciones fisiológicas específicas de dichas partes, podríamos resolver científicamente el problema mente-cuerpo, y recomendaba que empezáramos nuestra búsqueda centrándonos en un solo correlato de la conciencia, los procesos cerebrales que distinguen la visión consciente de la inconsciente. Cabe suponer que, a partir de ahí, podríamos extrapolar el resto de la conciencia. Se trata de una suposición bastante razonable, pues sin duda debe de haber un correlato neuronal de la experiencia visual, ¿no?

Como hemos visto, si bien solo la visión cortical es consciente, la corteza también puede procesar los estímulos visuales de manera inconsciente, y lo mismo pueden hacer los tubérculos cuadrigéminos superiores. En consecuencia, según Crick, el problema de la conciencia visual se reduce a la sencilla cuestión de preguntar qué es lo que tiene lugar en la corteza visual cuando está procesando información conscientemente y qué es lo que no tiene lugar cuando lo hace de manera inconsciente, algo que tampoco tiene lugar en los tubérculos cuadrigéminos superiores. Por motivos que he expuesto en páginas anteriores, creo que, en lo tocante al aspecto anatómico, Crick empezó con mal pie; debería haberse centrado en el tronco encefálico, más que en la corteza, y en el afecto más que en la visión. No obstante, Chalmers tenía una reserva de más peso.

El enfoque de Crick —que llegó a ser el dominante en la neurociencia cognitiva— elide lo que Chalmers llama la parte «difícil» del problema mente-cuerpo. Aislar los correlatos neuronales de la conciencia es la parte «fácil». El citado enfoque se limita a identificar los procesos cerebrales específicos que se correlacionan con la experiencia, pero no explica cómo la causan. Esa es la parte difícil del problema: ¿cómo y por qué las actividades neurofisiológicas producen la experiencia de la conciencia?[391] En otras palabras, ¿cómo la materia llega a ser mente? Según Chalmers, nosotros, los neurocientíficos, podemos ser

capaces de explicar el modo en que la información neuronal se procesa en el cerebro mientras tenemos experiencias visuales, pero eso no explica la manera en que esos procesos cerebrales se convierten en experiencias. En palabras de John Searle, otro prestigioso filósofo: «¿Cómo el cerebro es capaz de algo tan difícil como pasar de la electroquímica al sentimiento?».[392] Se trata de una pregunta igualmente desconcertante si se formula a la inversa: ¿de qué manera las cosas inmateriales como los pensamientos y los sentimientos (por ejemplo, tomar la decisión de preparar una taza de té) se convierten en acciones físicas; por ejemplo, preparar una taza de té?[393]

La extensión de ese vacío explicativo,[394] como lo llaman los filósofos, la ilustra bien el «argumento de conocimiento», que consiste más o menos en lo siguiente.[395] Imaginemos a una neurocientífica llamada Mary, ciega de nacimiento, que sabe todo lo que hay que saber sobre los correlatos neuronales de la visión. Aunque es capaz de explicar todos los hechos físicos del procesamiento de la información visual —incluidos los aspectos de nivel celular, como el impacto de las ondas luminosas en los conos y bastones fotosensibles, la manera en que esas ondas se convierten en impulsos nerviosos que se propagan hacia la corteza por el cuerpo lateral geniculado, donde los siguen procesando columnas bien ordenadas de neuronas, organizadas en grandes cantidades para formar una variedad de módulos de procesamiento de información extendidos por todo el manto cortical, del que se conocen bien muchos flujos de procesamiento especializados—, Mary no sabe cómo es experimentar la visión. Ciega de nacimiento, no sabría nada de las cualidades experimentadas de lo rojo y lo azul —por ejemplo—, que, a fin de cuentas, son la verdadera materia de la visión consciente. Y ello no solo es así porque Mary nunca ha experimentado tales cualidades, sino también porque nada en su conocimiento anatómico y fisiológico de los correlatos neuronales de la visión explica cómo es ver. Si Mary adquiriese de repente el don de la vista, aprendería algo totalmente nuevo sobre la visión, algo para lo que su comprensión mecanicista no la ha preparado en absoluto. Por tanto, los hechos físicos no explican por qué ni cómo hay algo que es como ver; solo explican por qué y cómo el cerebro descodifica la información visual, es decir, cómo ve el cerebro, no cómo vemos nosotros. Esta supuesta irreductibilidad a la base física de lo que los filósofos llaman qualia —el «algo que es como» de la experiencia subjetiva— es el problema difícil (según Chalmers, el «principal misterio de la conciencia»).[396]

Esta irreductibilidad percibida ha llevado a grandes mentes de todas las épocas, entre ellas (y no es poco) la del médico y filósofo John Locke, a concluir que las experiencias conscientes no son parte del

universo físico.[397] Dado, no obstante, que está claro que los qualia de experiencia existen, esos pensadores los relegaron a alguna dimensión no física de la realidad, una realidad que ellos (mejor dicho, muchos de ellos) llamaron «epifenoménica».[398] Fue Locke quien señaló, en su «argumento del espectro invertido», que desde un punto de vista lógico es posible que alguien experimente la cualidad de azul como yo experimento la de rojo y, aun así, la llame «rojo» como hago yo a pesar de que lo que ve conscientemente esa persona es lo que yo llamaría «azul». La conclusión del argumento es que no habría ninguna diferencia si esa persona también lo llamara «azul»: esos dos extremos del espectro de colores podrían intercambiarse sin problema en relación con los qualia que hubiese experimentado. En consecuencia, según Locke, la conciencia no se explica por sus correlatos físicos; la relación entre el mecanismo físico (en este caso, las longitudes de onda relativas de la luz) y los qualia psicológicos no es causal.

Así, la conciencia pasa a ser algo que simplemente discurre junto a la cadena de los sucesos físicos que tienen lugar en el cerebro —una suerte de subproducto— sin afectar a la estructura causal. Locke escribía en el siglo XVII, pero su punto de vista no es arcaico; antes bien, está muy extendido, y no solo entre filósofos. Como dijeron no hace mucho dos respetados científicos cognitivos: «La conciencia personal es análoga al arcoíris que acompaña a los procesos físicos en la atmósfera pero sin influir en ellos».[399]

Si las experiencias conscientes no desempeñan papel alguno en los mecanismos del mundo físico, ¿por qué existen? ¿Qué añade la conciencia al proceso de información que de todos modos se lleva a cabo de manera inconsciente? ¿Para qué sirve percatarse de los procesos cerebrales si la conciencia no influye en ellos?[400]

Lo anterior no se aplica solo a la visión. Uno de los casos más convincentes de evidencia experimental a favor de que la conciencia es un mero epifenómeno de los procesos cerebrales es la observación de Benjamin Libet según la cual a nuestra decisión subjetiva de iniciar un movimiento la preceden —pongamos, unos trescientos milisegundos antes— ondas cerebrales mensurables que anuncian el comienzo del movimiento que creemos haber iniciado.[401] En otras palabras, el comienzo físico del movimiento (en el cerebro) empieza antes de que decidamos conscientemente movernos; no somos de verdad «nosotros» quienes lo iniciamos. Está ampliamente aceptado que este hallazgo demuestra que la elección consciente —el «libre albedrío»— es una ilusión. Si el libre albedrío no existe, ¿qué queda, entonces, de la conciencia? Como hemos visto, Crick respondió

diciendo que somos nosotros —nuestras alegrías y nuestras penas— los que «en realidad» no existimos, que bastaría con tener una adecuada descripción física de tales qualia para poder darlos por explicados.

Nunca he podido aceptar ni el argumento de que los qualia conscientes existen en algún universo paralelo ni el que afirma que no existen en absoluto. Y ustedes tampoco deberían, porque ustedes son su conciencia. Afirmar que somos como un arcoíris que no influye en nuestro cuerpo físico es a todas luces absurdo, y también lo es afirmar que en realidad solo somos el comportamiento de las neuronas y que, por lo tanto, no existimos realmente. Cada momento de nuestra experiencia contradice esas afirmaciones.

La afirmación de que no existimos también se opone por completo a una de las más célebres conclusiones de todo el pensamiento filosófico. Después de una vida entera dedicada a reflexionar sobre aquello de lo que podía estar absolutamente seguro, René Descartes (en su «filosofía de la duda») llegó a la famosa conclusión de que lo único de lo que no tenemos que dudar es del hecho de que existimos. «Pienso, luego existo»; en otras palabras, tenemos experiencias, luego existimos.

Ese es el problema difícil de Chalmers. Dado que está claro que existe un yo que experimenta, ¿cómo lo acomodamos dentro de nuestra concepción física del universo?

No soy filósofo. Como ya han visto, mi interés por la neurología de la conciencia surgió y se desarrolló al margen de la literatura filosófica, y confieso que gran parte de lo que leí me resultó más bien desconcertante. No obstante, en el presente libro he intentado dar una respuesta científica natural al «problema difícil». No pretendo decir que he aclarado todos los misterios de la metafísica de la conciencia. Como veremos, siguen sin respuesta preguntas muy difíciles, pero sí creo que he demostrado el modo en que los procesos naturales, al desarrollarse según sus diversas necesidades, pueden (a lo largo del tiempo evolutivo) generar algo muy parecido a nuestros mundos de experiencia privados.

¿Cómo pesa ese hecho en el problema tal como lo formula Chalmers? ¿Cómo salva la brecha explicativa entre los mundos interior y exterior la explicación que he ofrecido?

Empecemos por una observación básica: las leyes que rigen las funciones mentales como la percepción, la memoria y el lenguaje son abstraídas de los datos subjetivos y objetivos. Un ejemplo sería la ley de Ribot, una abstracción científica relativa a la memoria a largo plazo, que explica el hecho observable de que nuestra memoria experimentada internamente de lo que ocurrió hace diez años está más y mejor consolidada que la memoria de lo que ocurrió hace diez minutos. Esa es la razón por la cual la gente de edad tiende a olvidar los hechos recientes y no los remotos. Lo mismo puede decirse de las huellas mnémicas fisiológicas observadas externamente: los recuerdos de diez años están más consolidados que los de hechos ocurridos diez minutos antes (exactamente al mismo nivel).[402] Así pues, el gradiente temporal de la ley de Ribot no es psicológico ni fisiológico, es ambas cosas. Consideremos asimismo la ley de Miller, una abstracción científica acerca de la memoria a corto plazo, que explica la limitación experimentada de nuestra capacidad para retener cosas en la mente; en un momento dado solo podemos retener siete unidades de información (dos más, dos menos). La duración de los recuerdos a corto plazo también puede medirse; lo común es que duren entre quince y treinta segundos.[403] Esa misma limitación de la capacidad puede observarse a nivel fisiológico, porque implica agotamiento de los neurotransmisores.[404] Por lo tanto, la ley de Miller, como la de Ribot, es a la vez psicológica y fisiológica.

Leyes como las citadas, que en principio se pueden cuantificar y, en consecuencia, expresarse en términos matemáticos, explican las manifestaciones duales de una función abstraída llamada «memoria». No cabe duda de que lo mismo ocurre con la «conciencia». Si los fenómenos de la conciencia son cosas naturales (¿qué si no?), también deben de ser reducibles a leyes.

No es una conclusión radical o idiosincrásica. Toda la ciencia cognitiva, que ha dominado mi campo de estudio desde finales del siglo XX, se basa en ella. Recordemos que en el capítulo 1 decía:

¿Qué es el procesamiento de la información? Lo desarrollaré a fondo más adelante, pero lo que nos interesa más ahora es que puede realizarse con todo tipo de equipos físicos. Esto arroja nueva luz sobre la naturaleza física de la mente, porque sugiere que la mente (interpretada como procesamiento de la información) es, más que una estructura, una función. Visto así, las funciones software de la mente las realizan las estructuras hardware del cerebro, pero pueden realizarlas igual de bien otros sustratos, como los ordenadores. Así,

unos y otros, cerebros y ordenadores, efectúan funciones de memoria (codifican y almacenan información) y funciones perceptivas (clasifican patrones de información entrante comparándolos con la información almacenada), además de funciones ejecutivas (ejecutan decisiones sobre qué hacer en respuesta a dicha información).

Sin embargo, de inmediato añadí:

Ahí radica la fuerza de lo que se acabó llamando el «enfoque funcionalista», pero también su debilidad. Si los ordenadores —que en teoría no son seres sintientes— pueden efectuar las mismas funciones, ¿de verdad podemos reducir la mente a un mero procesamiento de la información?

Estas cuestiones nos llevan al centro del problema difícil. Según Chalmers, el problema difícil de la conciencia no se resuelve con «explicaciones funcionalistas»:

Los problemas sencillos lo son precisamente porque se refieren a la explicación de las capacidades y funciones cognitivas. Para explicar una función cognitiva solo necesitamos un mecanismo capaz de desempeñar dicha función. Los métodos de la ciencia cognitiva se adecuan muy bien a esa clase de explicación, y, por tanto, también a los problemas sencillos de la conciencia. En cambio, el problema difícil es difícil precisamente porque no es un problema relativo a la ejecución de las funciones. El problema persiste incluso cuando se explica la ejecución de todas las funciones relevantes […]. Lo que hace que el problema difícil sea difícil y casi único es el hecho de que va más allá de los problemas relativos a la ejecución de las funciones. Para comprenderlo, tengamos en cuenta que incluso tras haber explicado la ejecución de todas las funciones cognitivas y conductuales en las inmediaciones de la experiencia, […] puede seguir habiendo otra pregunta sin respuesta: ¿Por qué la ejecución de esas funciones va acompañada de experiencia? Una explicación simple de las funciones deja abierta la siguiente pregunta […] ¿Por qué no tiene lugar todo el procesamiento de información «en la oscuridad», sin sensaciones interiores?[405]

En el capítulo 4 esbozo los intentos fallidos de especificar un mecanismo capaz de desempeñar la función de la conciencia. La respuesta de Chalmers al primero de dichos intentos transmite el tenor de su actitud respecto de todos ellos:

Crick y Koch sugieren que las oscilaciones [de las ondas gamma sincronizadas] son los correlatos neuronales de la experiencia. Se trata de una afirmación discutible […], pero incluso si se acepta, sigue sin responderse a la cuestión de la explicación: ¿por qué las oscilaciones dan lugar a experiencia? Solo se encuentra una conexión explicativa en el papel que desempeñan las oscilaciones en la vinculación y el almacenamiento, pero nunca se aborda la razón de que esa vinculación y almacenamiento deban ir acompañados de experiencia, y, si no sabemos por qué deberían dar lugar a la experiencia, de nada sirve buscar la explicación en las oscilaciones. A la inversa, si supiéramos por qué la vinculación y el almacenamiento dan lugar a experiencia, los detalles neurofisiológicos no serían más que la guinda del pastel. La teoría de Crick y Koch saca ventaja suponiendo una conexión entre vinculación y experiencia; por tanto, no sirve para explicar esa conexión.[406]

Veamos ahora la crítica de Chalmers a la segunda explicación funcionalista más importante de la conciencia, la teoría del «espacio de trabajo global» de Newman y Baars (1993):

Se podría suponer, de acuerdo con Baars, que los contenidos de experiencia son precisamente los contenidos del espacio de trabajo, pero incluso si es así, nada interno a la teoría explica por qué se experimenta la información dentro del espacio de trabajo global. Lo mejor que la teoría puede hacer es decir que la información se experimenta porque es globalmente accesible. Sin embargo, ahora la pregunta reaparece reformulada: ¿por qué la accesibilidad global da lugar a la experiencia consciente? Como siempre, esta pregunta puente sigue sin respuesta.[407]

Chalmers concluye:

Los métodos explicativos habituales de la ciencia cognitiva y de la neurociencia no son suficientes. Estos métodos se han desarrollado precisamente para explicar la ejecución de las funciones cognitivas, cosa que hacen muy bien. Sin embargo, en su estado actual, solamente están equipados para explicar la ejecución de funciones. A la hora de abordar el problema difícil, el enfoque estándar no tiene nada que decir [...]. Para explicar la experiencia hace falta un nuevo enfoque. [408]

Tengo que ser claro respecto de estas palabras de Chalmers. Dice que los métodos de la ciencia cognitiva solo pueden explicar la ejecución de las funciones cognitivas, y que lo hacen bien. Con «funciones cognitivas» se refiere a cosas como la percepción, la memoria y el lenguaje, o sea, dice que la conciencia no es una función cognitiva. ¿Por qué lo dice? La respuesta es: porque no habla de una función abstraída llamada «conciencia», sino más bien de la experiencia de la conciencia, es decir, habla de cómo es percibir o recordar.

El problema difícil sería trivial si en su totalidad se redujese al hecho de que nuestra experiencia individual no es igual a la experiencia humana en general.[409] Si ese fuera el problema difícil, lo único que tendríamos que hacer para resolverlo sería tomar las experiencias de testigo único de muchos individuos, promediarlas, encontrar el denominador común y explicar eso en términos de funciones. Algo parecido hacen sin cesar los psicólogos, y eso fue lo que hice cuando investigué la fabulación: empezar por las experiencias subjetivas de un individuo, el señor S., generalizarlas estudiando las experiencias equivalentes de otros pacientes como él y luego abstraer el común denominador. El enfoque reveló un principio funcional relativo a la fabulación, a saber, que consigue que los pacientes se sientan mejor; la fabulación sirve a una función «anhelante».

Con todo, Chalmers no se limita a decir que los fenómenos subjetivos están conectados con un único punto de vista; lo que hace es preguntar por qué una función cualquiera debería estar acompañada de experiencia. En consecuencia, escribe:

¿Por qué cuando las formas de ondas electromagnéticas impactan en una retina y son discriminadas y categorizadas por un sistema visual,

esa discriminación y esa categorización [función] se experimentan como una sensación de rojo vivo? Sabemos que la experiencia consciente surge cuando esas funciones se ejecutan, pero el principal misterio es el hecho mismo de que surja. Hay una brecha explicativa (término que debemos a Levine, 1983) entre las funciones y la experiencia, y necesitamos un puente explicativo para salvarla. Limitarse a describir las funciones equivale a quedarse a un lado de la brecha; los materiales para el puente hay que buscarlos en otra parte. [410]

Merece la pena explicar bien lo que ese «puente explicativo» debe conseguir, no vaya a ser que movamos la portería. En un artículo publicado poco después de que Friston y yo presentáramos el nuestro (2018), tres colegas checos anunciaron que la solución al «problema difícil» era inminente y predijeron que provendría del principio de la energía libre:

Entre los filósofos y los neurocientíficos no hay una comprensión clara ni un consenso general acerca de la función de la conciencia. Esa es una de las razones principales por las que la conciencia sigue suponiendo un problema tan difícil, algo que hunde sus raíces en el hecho de que nunca ha habido una función articulada de conciencia basada en —y sostenida por— alguna teoría del cerebro unificadora. Esa teoría unificadora empieza a emerger ahora, en forma del marco de codificación predictiva [...] basado en las ideas de Hermann von Helmholtz de que el cerebro es ante todo una máquina de inferencia predictiva.[411]

No obstante, estos autores añadían una advertencia: «Por desgracia, el problema difícil de la conciencia no se resolverá nunca por completo porque siempre tendrá partidarios muy devotos». Y continuaron diciendo:

Creemos que apenas puede esperarse que los defensores más fieles del problema difícil lleguen a estar plenamente satisfechos con [cualesquiera] conclusiones de la ciencia empírica, puesto que el argumento central del problema difícil se dirige ante todo a los empeños de la ciencia empírica.

Como no podía ser de otra manera, cuando Friston y yo recibimos las primeras reseñas de colegas del Journal of Consciousness Studies, eso fue exactamente lo que uno de ellos dijo sobre nuestro artículo: «El problema difícil (según Chalmers) es un problema metafísico, y en cuanto tal no está abierto para ser "resuelto"». (Después, Friston me escribió: «Tengo la impresión de que el problema difícil no está ahí para resolverlo; está para venerarlo»).[412]

Para ser justos con Chalmers, cabe decir que no es responsable de los prejuicios filosóficos que preocupaban a nuestros colegas checos. Veamos, por ejemplo, la última frase de su artículo de 1995: «El problema difícil es un problema difícil, pero no hay razón alguna para creer que nunca se resolverá».

Así pues, hablemos claramente sobre lo que debe lograr el puente explicativo que Chalmers pide. Citaré aquí sus propios criterios. En su derribo de la primera teoría funcionalista, Chalmers escribió que nunca se había abordado por qué la vinculación y el almacenamiento deberían ir acompañados de experiencia. Así pues, al evaluar si la brecha explicativa se ha salvado gracias a una teoría funcionalista de la conciencia hay que preguntarse: ¿se ha abordado la cuestión relativa a por qué la función XYZ debería ir acompañada de experiencia? En lo que atañe a la segunda teoría rechazada por Chalmers, añadió: «¿Por qué la accesibilidad global debería dar lugar a experiencia consciente? Como siempre, esta pregunta puente sigue sin respuesta». De ahí que al evaluar cualquier otra teoría haya que preguntarse: ¿se ha respondido ya a la cuestión relativa a por qué la función XYZ debería dar lugar a experiencia consciente? Y al parecer, eso es todo.

Pero, si eso es todo, ¿por qué Chalmers añade que «limitarse a describir las funciones [de la experiencia] equivale a quedarse en un lado de la brecha; los materiales para el puente hay que buscarlos en otra parte»? ¿Por qué una explicación de las funciones no sirve para salvar la brecha? Releyendo lo que Chalmers dice a este respecto, tal como he citado antes, es evidente que confunde dos clases distintas de brecha explicativa. Por tanto, si vamos a intentar salvarla, hemos de establecer con claridad de qué brecha hablamos.

La primera es una brecha explicativa entre señales propagadas desde la retina y sensaciones de rojo vivo, es decir, entre sucesos fisiológicos

y sucesos psicológicos. Me veo obligado a señalar que ambos tipos de sucesos son, de hecho, experimentables. Es posible experimentar sucesos fisiológicos tales como la observación de señales propagadas desde la retina con la misma facilidad con la que es posible experimentar sucesos psicológicos como la sensación de rojo vivo. Ninguno de ellos puede ser explicado —y aun menos darse por explicado— por el otro. Son dos maneras de observar lo mismo. Cuando (introspectivamente) experimento que existo, ¿es la cosa mental que existe distinta del Mark Solms corporal que veo en el espejo? ¿Y el Mark Solms del espejo es algo distinto del Mark Solms anatómico que veo en una imagen por resonancia magnética? No comprendo por qué son tantos (filósofos incluidos) los que hablan del cerebro como si en cierto modo estuviera exento de la realidad tal como la experimentamos.[413] Solo se explica porque cuando dicen «el cerebro» en realidad se refieren a otra cosa. No hablan del cerebro que vemos y tocamos, sino de algo abstraído de nuestra experiencia del cerebro, algo parecido a un sistema funcional.

Por tanto, la primera brecha explicativa está situada entre dos clases distintas de experiencia asociadas a dos perspectivas observacionales diferentes,[414] algo análogo a oír truenos con los oídos y ver relámpagos con los ojos. Nadie dice: «Es un hecho sabido que el trueno viene después del relámpago, pero carecemos de una buena explicación que nos diga por qué y cómo». Y no se dice porque nadie cree que el relámpago produzca el trueno de la misma manera en que el hígado produce la bilis; se acepta que se trata de dos manifestaciones del mismo elemento subyacente,[415] algo aplicable también a las distintas maneras de experimentar el procesamiento de información visual, a saber, desde el exterior o desde el interior. Desde fuera (si uno es un científico con el equipo adecuado), se ven señales propagadas desde la retina;[416] desde dentro se ve rojo vivo.

La segunda brecha explicativa está situada entre experiencias (de ambas clases) y sus causas subyacentes. Es la brecha entre cosas que podemos experimentar, tales como rojos vivos e imágenes optogenéticas de neuronas activadas, y cosas que no podemos experimentar, como campos cuánticos en sí. Se trata, en pocas palabras, de una brecha entre las perspectivas de la primera y la tercera personas. Adoptar la perspectiva de una tercera persona sobre mi propia experiencia equivale a abstraerme a mí de la experiencia y dejar de experimentarla. Esta perspectiva no concierne ni al cerebro tal como se ve ni a la mente tal como se siente, sino más bien a las fuerzas que explican el porqué y el cómo de esa apariencia y esa sensación, y esa es la perspectiva que he adoptado en este libro.

No sé si ha quedado claro que Chalmers, en la cita antes reproducida, se refiere a ambas brechas a la vez, algo que puede llevar a confusión. Chalmers toma la brecha entre las dos clases de experiencia (observaciones exterospectivas frente a interospectivas) y la confunde con la brecha entre experiencia en general y sus mecanismos funcionales subyacentes. Y esto hace que el abismo que hay que salvar sea mucho más grande.

Tengo la impresión de que la mayoría supone que lo que Chalmers llama «función» es sinónimo de lo que llama «lo físico», y que «experiencia» es, por tanto, solo y siempre algo no físico. No obstante, si «lo físico» significa el cuerpo observable y sus órganos, incluido el cerebro, entonces es objeto de experiencia no menos que lo psicológico.[417] Además, las experiencias psicológicas, si se abstraen, revelan el mecanismo funcional de tales experiencias y dan lugar a leyes psicológicas.

Lo mismo se aplica a las leyes fisiológicas: también se abstraen de la experiencia, de los datos fisiológicos observables. Estas dos clases de leyes, al estar hechas ambas del mismo material abstraído (explicativo), o sea, las funciones, no son tan difíciles de reducir una a la otra como los dos tipos radicalmente distintos del acto de experimentar. Así pues, la ley de Miller y la ley de Ribot son tanto psicológicas como fisiológicas, y por ese motivo pueden reducirse a ecuaciones unificadoras. El no poder hacer explícitos los pasos intermedios entre datos basados en la experiencia y mecanismos explicativos exagera el problema difícil y hace que parezca más difícil de lo que en realidad es. No hay que intentar imaginar el modo en que una clase de experiencia podría producir otra clase de experiencia; basta con encontrar una teoría mecanicista que explique ambas series de fenómenos, de la modalidad de perspectiva que fuere. Después ya probaremos las predicciones a las que dé lugar.

Chalmers admite que el problema difícil se puede resolver. De hecho, establece tres principios en los que debería basarse la solución. Antes de exponer aquí esos principios (y uno de ellos tenemos que dejarlo para el capítulo 12), debo aclarar que Chalmers cree que no podremos resolver nunca el problema difícil si intentamos hacerlo reduciendo «la experiencia» a lo que él llama «procesos físicos». Chalmers escribe:

Ya estamos en condiciones de comprender algunos hechos claves

sobre la relación entre procesos físicos y experiencia, y acerca de las regularidades que los conectan. Una vez que dejamos de lado la explicación reductiva, podemos poner esas verdades sobre la mesa de modo tal que desempeñen su apropiado papel de piezas iniciales en una teoría no reductiva de la conciencia, y como limitaciones de las leyes básicas que constituyen una teoría definitiva.[418]

Personalmente opino que una explicación «no reductiva» es algo bueno si significa que podemos renunciar a la imposible tarea de reducir los fenómenos psicológicos a fenómenos fisiológicos, o viceversa. Los fenómenos psicológicos no se pueden reducir a fisiológicos, no más que el relámpago puede reducirse al trueno. El relámpago no es la causa del trueno; los dos fenómenos se correlacionan entre sí. Ese es el problema fácil. En consecuencia, debemos reducir ambos fenómenos a sus respectivos mecanismos, de modo tal que podamos reducir dichos mecanismos a un denominador común sin violar las leyes de la física. Ese es el problema difícil.

No obstante, cuando dice explicación «no reductiva», Chalmers parece querer decir otra cosa. Para él significa que no podemos reducir fenómenos psicológicos experimentados a leyes funcionales, punto.

Y prosigue:

Una teoría no reductiva de la conciencia consistirá en un número de principios psicofísicos que conecten las propiedades de los procesos físicos con las propiedades de la experiencia. Podemos pensar que esos principios sintetizan la manera en que la experiencia surge de lo físico. En última instancia, dichos principios deberían decirnos qué clase de sistemas físicos tendrán experiencias asociadas, y, para los sistemas que las tienen, deberían decirnos qué clase de propiedades físicas son relevantes para que emerja experiencia y exactamente qué clase de experiencia deberíamos esperar que produzca un sistema físico dado. Tarea difícil.

Ese es el primer principio de Chalmers, que él llama «principio de coherencia estructural», pues implica coherencia estructural entre «las propiedades de la experiencia» y las «propiedades de los procesos físicos». Aquí parece referirse a dos clases de fenómenos que, en mi opinión —como en la suya—, no pueden reducirse directamente el

uno al otro. Sin embargo, Chalmers también habla de «clase[s] de sistemas físicos» y sus «experiencias asociadas», lo cual parece referirse a la relación entre sistemas funcionales en general y experiencias en general (por lo que las segundas «emergen de» las primeras o las primeras «producen» las últimas). Para Chalmers, ese es el motivo por el cual el principio de coherencia estructural sintetiza la manera en que «la experiencia surge de lo físico».

Una vez más, Chalmers confunde los fenómenos físicos y las causas «físicas» (es decir, funcionales). A menos que desglosemos los pasos intermedios, no podremos sino concluir que la experiencia no surge de lo físico. El mismo significado confuso de «lo físico» puede verse en las típicas teorías funcionalistas que he mencionado antes, las cuales buscaban la llave de la experiencia psicológica en los mecanismos fisiológicos X, Y o Z: la sincronía de las oscilaciones gamma; la vinculación y almacenamiento de la actividad sensorial y frontoparietal integrada; la activación de la corteza por los núcleos intralaminares del tálamo; los «bucles reentrantes» talamocorticales, *etc.* Si echamos mano de la analogía entre el trueno y el relámpago, esa refundición requiere de nosotros que expliquemos los fenómenos auditivos a través de los mecanismos funcionales de la visión. Solo en este sentido se justifica la afirmación de Chalmers de que «limitarse a describir las funciones equivale a quedarse en un lado de la brecha; los materiales para el puente hay que buscarlos en otra parte».

Si queremos explicar la psicología en relación con la fisiología, tenemos que abstraernos de los fenómenos observados de ambas clases (esto es, tenemos que inferir mecanismos funcionales de ambas clases) y luego abstraernos de las dos series de abstracciones para ver el denominador común unificador. En el proceso debemos situarnos a igual distancia de ambas (es decir, debemos inferir mecanismos lo bastante profundos para dar cuenta de las funciones tanto de la psicología como de la fisiología). Solo entonces podremos conciliar entre sí los fenómenos con sus mecanismos subyacentes.

No quisiera sonar demasiado manido a este respecto. La ciencia real procede de maneras menos ordenadas. Empezamos con una comprensión o un descubrimiento dentro de la cascada de la causalidad y luego llenamos los espacios en blanco. Sin embargo, pasando por alto los pasos intermedios llegamos a la exigencia de soluciones «no reductivas» al problema difícil que son imposibles de alcanzar. Recordemos las primeras preguntas de Chalmers: ¿por qué el procesamiento físico debería dar lugar a una rica vida interior?; ¿cómo y por qué las actividades neurofisiológicas producen la experiencia de la conciencia? Por la manera en que están formuladas,

son preguntas que no pueden contestarse no reductivamente.

Es comprensible, por tanto, que Chalmers llegue a la conclusión de que una solución reductiva per se es imposible:

Personalmente, sugiero que una teoría de la conciencia debería tomar la experiencia como fundamental. Sabemos que una teoría de la conciencia requiere que se añada algo fundamental para nuestra ontología, pues todo en la teoría física es compatible con la falta de conciencia. […] Una teoría no reductiva de la experiencia especificará los principios básicos que nos digan el modo en que la experiencia depende de las características físicas del mundo. Tales principios psicofísicos no interferirán en las leyes físicas, puesto que parece que las leyes físicas ya forman un sistema cerrado. Antes bien, serán un suplemento de una teoría física. Una teoría física da una teoría de los procesos físicos, y una teoría psicofísica nos dice cómo esos procesos dan lugar a la conciencia. Sabemos que la experiencia depende de los procesos físicos, pero sabemos también que esa dependencia no puede derivarse solo de las leyes físicas. Los nuevos principios básicos postulados por una teoría no reductiva nos brindan el ingrediente extra que necesitamos para construir un puente explicativo. […] Esta posición puede considerarse una variedad de dualismo, pues postula propiedades básicas más allá de las propiedades que invoca la física. Sin embargo, es una explicación inocente de dualismo.

Y así, al final, Chalmers dice algo que espanta. Puesto que cree que la «experiencia» no puede reducirse a «lo físico» (en el sentido en que él emplea la palabra, con los significados «fisiológico» y «funcional» a la vez), se ve obligado a concluir que la experiencia no forma parte del universo físico conocido, una conclusión que suena a dualismo puro, de la clase no tan inocente que nos legaron Descartes y Locke.[419] He ahí el motivo por el que Chalmers afirma que la conciencia requiere que se añada «algo fundamental para nuestra ontología»; por ejemplo, algo no físico «muy por encima de las propiedades evocadas por la física» o algo para «complementar» las leyes físicas. Espero dejar claro por qué no estoy de acuerdo con él en este punto. A la luz de todo lo que he expuesto hasta aquí, no puedo estar de acuerdo en que «todo en la teoría física es compatible con la falta de conciencia». Pero hay más cosas… Tras declararse en contra de reducir la «experiencia» a «lo físico», Chalmers prosigue diciendo (como para ilustrar la confusión) que, no obstante, son aspectos duales de otra

cosa. Esa «otra cosa», según él, es información; que, de acuerdo, no es algo físico en sentido fisiológico, pero que, en mi opinión, es algo físico en el sentido funcional de la mecánica estadística. Este movimiento introduce el segundo principio de Chalmers, al que llama «principio del doble aspecto» y al que considera más básico que el principio de coherencia estructural:

El principio básico que sugiero implica en esencia la noción de información, y entiendo información más o menos en el sentido de Shannon (1948). Allí donde hay información, hay estados de información incrustados en un espacio de información. [...] Un espacio de información es un objeto abstracto, pero, según Shannon, podemos considerar que la información está físicamente incorporada cuando hay un espacio de claros estados físicos y las diferencias entre ellos pueden transmitirse por alguna ruta causal [...]. El principio del doble aspecto deriva de la observación de que hay un isomorfismo directo entre ciertos espacios de información físicamente incorporados y ciertos espacios de información fenoménicos (o experienciales). A partir de la misma clase de observaciones relativas al principio de coherencia estructural, podemos advertir que las diferencias entre estados fenoménicos tienen una estructura que se corresponde directamente con las diferencias incrustadas en procesos físicos; en particular, con aquellas diferencias que hacen que sean distintas ciertas rutas causales implicadas en la disponibilidad y el control globales. Es decir, podemos encontrar el mismo espacio de información abstracto incrustado en el procesamiento físico y la experiencia consciente. Esto conduce a una hipótesis natural, a saber, que la información (o, por lo menos, alguna información) tiene dos aspectos básicos, uno físico y uno fenoménico. Tiene la condición de un principio básico que podría subyacer a —y explicar— la emergencia de la experiencia a partir de lo físico. La experiencia surge en virtud de su condición como un aspecto de la información, cuando el otro aspecto se encuentra incorporado en el procesamiento físico.

Obsérvese que aquí Chalmers describe dos veces la información como algo «abstracto». En consecuencia, una vez diferenciados los dos significados de «lo físico», queda claro que la información tiene el mismo estatus ontológico en relación con sus aspectos fisiológico y psicológico, del mismo modo que la electricidad lo tiene en relación con el relámpago y el trueno. El relámpago y el trueno son aspectos duales de la descarga eléctrica como las señales propagadas desde la

retina y la sensación de rojo vivo son aspectos duales del procesamiento de información; son distintas respuestas suscitadas por el equipo, diferentes manifestaciones fenoménicas de un proceso causal unitario. Y ese proceso causal es físico en el sentido explicativo.

En este contexto, Chalmers cita incluso al físico John Wheeler. Como he explicado antes, Wheeler introdujo una interpretación «participativa» de la mecánica cuántica, en cuyos términos el universo observado se manifiesta en respuesta a las preguntas que se le formulan sobre él. Esa es la razón por la cual la misma cosa puede adoptar formas complementarias, tal como ocurre con las ondas y las partículas. Según la interpretación de Wheeler, la forma fenoménica que adopta el universo experimentado depende del modo en que es observado o medido, es decir, de nuestra perspectiva del mismo. Por ejemplo, un recuerdo se puede experimentar como una reminiscencia o como una huella neuronal activada, según el equipo que utilicemos para observarlos. La forma que adquiera está en el ojo del espectador.

Sin embargo, Chalmers no lo ve así. Para él, la complementariedad es inherente a la cosa observada, no al observador ni tampoco a la acción de observar:

Las leyes de la física pueden plantearse en términos de información, postulando distintos estados que dan lugar a distintos efectos sin decir en realidad cuáles son. Lo único que importa es su posición en un espacio de información. De ser así, la información es candidata natural para desempeñar un papel también en una teoría fundamental de la conciencia. Nos vemos abocados a una concepción del mundo en la que la información es de verdad fundamental y en la que tiene dos aspectos básicos, correspondientes a los rasgos físicos y fenoménicos del mundo.[420]

Así pues, Chalmers piensa que los aspectos duales de «los rasgos fenoménicos del mundo» y «lo físico» están en la información en sí como su fuente), no en el equipo del observador participante.[421] Esto es como pensar que la dulzura es intrínseca a la estructura molecular de la glucosa.[422] Es posible que esta distinción no importe cuando lo observado es el observador; como cuando vemos nuestro cuerpo en un espejo.[423] Y es posible que al final no importe tanto que Chalmers confunda dos significados de «lo físico» (a saber, los niveles fenoménico-fisiológico y mecánico-funcional) porque de

todos modos reduce a ambos a «información».

Con todo, hay dos cosas que siguen representando un problema. Primero, para Chalmers las propiedades de la experiencia y las de lo físico no son reducibles a información. Ambas son inherentes a la información. Segundo, para él ambas son inherentes a toda la información, y por ese motivo en otra parte de su artículo afirma que las «propiedades fenoménicas son el aspecto interno de la información», que él contrasta con las propiedades físicas (supuestamente no fenoménicas), que son el «aspecto externo». He ahí el extraño paso al que le conduce su dualismo.

Chalmers concluye:

Esto podría responder a una preocupación sobre la relevancia causal de la experiencia; una preocupación natural, dada una escena en la que el dominio físico está causalmente cerrado y en la que la experiencia es un complemento de lo físico. El punto de vista de la información nos permite entender el modo en que la experiencia podría tener una especie de sutil relevancia en virtud de su condición de naturaleza intrínseca de lo físico.

Espero que el presente libro los haya convencido de que la relevancia causal de la experiencia para sistemas autoorganizados complejos como nosotros es de todo menos sutil.

De todos modos, aún tenemos que explicar por qué y cómo la experiencia surge de manera legítima de mecanismos físicos. Chalmers pregunta: «¿Por qué la ejecución de estas funciones [físicas] va acompañada de experiencia?». Como hizo Nagel antes que él, Chalmers define la experiencia recurriendo al «algo que es como». Para Chalmers, esa es la esencia de la subjetividad, «el aspecto interno de la información», que es, a su vez, «la naturaleza intrínseca de lo físico». Prosigue argumentando que la subjetividad no puede reducirse a algo no subjetivo. Para él, esa es la razón por la cual la experiencia subjetiva debe ser inherente a la información.

No obstante, esto parecería dar a entender que la cualidad experimentada de ser algo (de ser información) es una propiedad fundamental de todas las cosas. ¿Por qué Chalmers piensa que ser

experimentada es «el aspecto interno de la información» en general? Si sostenemos, como he hecho aquí, que la subjetividad solo es una perspectiva observacional, donde un sujeto es simplemente el ser de cierta clase de objeto, entonces podemos decir libremente que para muchos objetos —de hecho, para la mayoría de ellos— no hay «algo que es como» ser ellos. Cuando se las considera desde su propio punto de vista subjetivo, la gran mayoría de las cosas no posee el carácter de «algo que es como».

¿Hay algo que es como ser una célula única o una planta? ¿Y una piedra? ¿Hay algo que es ser como un ordenador o internet? ¿Y un termostato? Chalmers dice: «Es innegable que algunos organismos son sujetos de la experiencia». ¿Significa eso que solo los organismos (o solo algunos organismos) son sujetos de la experiencia? Chalmers admite que estamos ante cuestiones abiertas y que puede ocurrir que la experiencia solamente surja a cierto nivel de complejidad o con cierta clase de procesamiento de información:[424]

Una cuestión obvia consiste en saber si toda la información tiene un aspecto fenoménico. Una posibilidad es que necesitamos una restricción adicional a la teoría fundamental, que indique con precisión qué clase de información tiene un aspecto fenoménico. La otra posibilidad es que tal limitación no exista. Si no existe, entonces la experiencia está mucho más extendida de lo que podríamos haber creído, pues la información está en todas partes. Al principio parece contraintuitivo, pero, si reflexionamos, creo que la posición adquiere cierta verosimilitud y cierta elegancia. Allí donde hay un procesamiento sencillo de la información, hay experiencia sencilla, y allí donde hay un procesamiento complejo de la información, hay experiencia compleja. Un ratón tiene una estructura de procesamiento de información más simple que un ser humano, y, en consecuencia, su experiencia es más simple. ¿Es posible que un termostato, una estructura de procesamiento de información de máxima simplicidad, pueda tener una experiencia de máxima simplicidad? De hecho, si la experiencia es de verdad una propiedad fundamental, sería sorprendente que surja solo de vez en cuando; las propiedades más fundamentales están extendidas de manera más uniforme. En todo caso, se trata en gran medida de una cuestión abierta.

Si aceptamos la opinión de Chalmers de que la experiencia es una propiedad fundamental de todo, entonces no hay nada que explicar, y

si la experiencia está en todas partes y es eterna, entonces tampoco hay mucho que temer. Por el precio de una sola especulación desenfrenada, apartamos el problema difícil y nos concedemos la inmortalidad al quitarnos de encima el miedo a que nuestra existencia pueda depender de algo como si fuera un mal sueño.

Ojalá fuera así. Por desgracia, como hemos visto en capítulos anteriores, la experiencia solo surge si se dan ciertas condiciones especiales. Está ubicada en un lugar exacto entre los flujos de información que atraviesan nuestros seres penosamente frágiles. Desempeña tareas concretas y luego desaparece. A todas luces, no es más básica para la estructura de la realidad que cualquier otra variedad de respuesta suscitada por el equipo. Sintiéndolo mucho, si no aceptamos la especulación de Chalmers, entonces tenemos que responder a esta pregunta: ¿por qué hay algo que es como ser algunas cosas y no otras? Si no todos los objetos son sujetos, y no todos los sujetos son sintientes, ¿cómo es que a veces surgen sujetos sintientes?

A mí la propuesta expansiva de Chalmers sobre la información me intriga, pero pienso que es mucho más verosímil que necesitemos «una restricción adicional a la teoría fundamental que indique con precisión qué clase de información tiene un aspecto fenoménico». Definir dicha limitación ha sido el objetivo principal de este libro. Hemos visto que todo depende del objetivo y la finalidad del procesamiento de información. En lo relativo a la conciencia, eso significa minimizar la entropía. Pero es algo más complicado que eso: se necesita también una manta de Markov, pues conlleva minimizar nuestra propia entropía. Además, conlleva hacerlo a través de una miríada de parámetros categóricos en contextos imprevistos. Los hechos físicos que he esbozado en este libro revelan que la conciencia no es inherente a toda la información, sino más bien a cierta clase de procesamiento de la información: una forma compleja del tipo autoevidenciable.

Si solamente algunas cosas, o solo los organismos, o solo algunos organismos son sujetos de experiencia, entonces la conciencia no puede ser una propiedad fundamental del universo. No cabe duda de que hubo un amanecer de la vida; las pruebas de ello son abundantes. Y la vida surgió muchísimo después del big bang. Por lo tanto, tuvo que transcurrir muchísimo tiempo antes de que existiera la conciencia. Solo esta suposición —hubo un amanecer de la conciencia— nos obliga a encontrar una explicación física de ello. Si hubo un amanecer, debió de haber algo previo a la conciencia que la explique. La idea alternativa —que la conciencia precedió a la vida y al universo— no encaja en los hechos como estos parecen ser y encima recuerda

demasiado a la idea de Dios. A menos que invoquemos ideas así, la conciencia debió de surgir —y, por tanto, debe de ser parte de él— de un universo físico no consciente.

Esto se aplica incluso dentro de nosotros mismos. En capítulos anteriores hemos visto que la percepción y la cognición no van necesariamente acompañadas de conciencia. De hecho, la evidencia científica sugiere que la percepción y la cognición son en su mayoría inconscientes (a este respecto he citado los artículos de revisión clásicos de Kihlstrom y de Bargh y Chartrand). Tras examinar la evidencia, estos científicos llegaron a la conclusión de que «la mayor parte del tiempo» no somos conscientes de nuestros actos psicológicos.

En los resultados empíricos que estos autores revisaron había muchas cosas que merecen nuestra atención (sobre todo, la manera en que la intencionalidad inconsciente deriva del aprendizaje consciente a través de la experiencia),[425] pero aquí quisiera recordar solo el resultado: hay hechos que pasan y existen dentro de nosotros seamos o no conscientes de ellos, incluidos los sucesos psicológicos. Podríamos haber llegado a esta conclusión —obvia en cierto sentido— desde muy distintos puntos de partida, pero no tiene importancia. Si «la mayor parte de la vida psicológica momento a momento» discurre sin experiencia, ¿por qué no pueden los termostatos e internet ser inconscientes? Si el ser de los núcleos basales no declarativos del cerebro no es «intrínsecamente experiencial» (por usar el término de Chalmers), ¿cómo puede ser experiencial toda la información?

No cabe duda de que la subjetividad en general debe estar hecha de algo no experiencial. No se trata de una observación metafísica sobre «cosas en sí» contra «cosas tal como las experimentamos»; es un hecho empírico. La vida psicológica momento a momento tal como la experimento no es en absoluto la totalidad de mi vida psicológica. Esto pone la experiencia en perspectiva. Como nos enseñó Freud, no toda la vida psicológica es consciente. Y solo podemos comprender la naturaleza de la conciencia discerniendo el trabajo particular que lleva a cabo; es decir, su función.

Yo no hago más que preguntar: ¿qué añade la conciencia al procesamiento de información (al procesamiento de información que de todos modos se produce de manera inconsciente)? ¿Qué sentido tiene tomar conciencia de los procesos físicos si nuestra conciencia no influye en esos procesos? A mi entender, esta pregunta es idéntica a la principal de Chalmers: «¿Por qué la ejecución de esas funciones va

acompañada de experiencia? […] ¿Por qué no tiene lugar todo el procesamiento de información "en la oscuridad", sin sensaciones interiores?».

En mi opinión, la pregunta solo surgió porque Chalmers, siguiendo a Crick, buscó la función de la conciencia donde no debía. La forma fundamental de la conciencia no es algo cognitivo, como la visión, sino algo afectivo. En ese sentido —y solo en ese sentido— Chalmers acertó al dar a entender que la conciencia no es una función cognitiva: la función principal de la conciencia no es percibir, recordar o comprender, sino sentir.

¿Cómo puede la función de sentir tener lugar «en la oscuridad», sin ningún sentimiento? Es legítimo preguntar por qué la visión se acompaña de experiencia. Al igual que cualquier otro proceso cognitivo, la visión no requiere conciencia, pero las sensaciones y los sentimientos sí.[426] Es verdad que algunos teóricos afirman que las pulsiones, los afectos y las emociones no son necesariamente conscientes, y esto es así porque cada persona atribuye distinto significado a esas palabras, y es por ese motivo que he empleado los términos sensación y sentimiento: si existen cosas como las emociones inconscientes, sensación y sentimiento designan a sus homólogas conscientes. No puede haber una sensación que no sintamos; «sensación inconsciente» es un oxímoron.

Por ese motivo, en este libro he centrado los argumentos científicos en las sensaciones y los sentimientos. Con vistas a resolver el problema difícil de la conciencia, la ciencia tiene que discernir las leyes que rigen la función mental de «sentir». No estamos ante una simple cuestión de palabras. He reunido un número considerable de pruebas para demostrar que las sensaciones y los sentimientos son la forma fundacional de la conciencia, su requisito previo. Asimismo he explicado desde el punto de vista tanto fisiológico como mecanicista la diferencia entre necesidades sentidas y no sentidas, y he demostrado que los sentimientos tienen consecuencias concretas, algo que me permitía concluir, en el capítulo 9, que «la conciencia no es meramente una perspectiva subjetiva de la dinámica "real" de los sistemas autoorganizados, sino una función con poderes causales definidos propios».

La perplejidad que se percibe en las preguntas de Chalmers —«¿Por qué la ejecución de esas funciones va acompañada de experiencia?» y «¿Por qué no tiene lugar todo el procesamiento de información "en la oscuridad", sin sensaciones interiores?»— se disipa cuando tomamos conciencia de que la «experiencia» no es intrínseca a la visión ni al

procesamiento de la información en general. Es intrínseca solo a la forma concreta del procesamiento de información que genera sentimientos.

Esa es, para mí, la razón por la cual no hay algo que es como ser un termostato o internet. Ahí reside la enorme pertinencia de las preguntas «¿Por qué hay algo que es como ser algunas cosas y no otras?» y «¿Cómo es que a veces surgen sujetos sintientes?». A estas preguntas hay que responder en términos mecanicistas. ¿Cuál es la función de sentir? ¿Puede el sentimiento ejecutar su función sin experiencia? Si Mary, la neurocientífica invidente, fuera una neurocientífica afectiva, ¿acaso no explicaría (de hecho, predeciría) su conocimiento de todo lo que hay que saber sobre la función de sentir por qué se siente como algo?[427] Lo mismo puede afirmarse del espectro invertido de Locke. ¿Es posible, desde un punto de vista lógico, que alguien experimente como atrozmente doloroso todo lo que yo experimento como exquisito y que eso no conlleve ninguna diferencia? Sin duda no, porque las sensaciones y los sentimientos hacen algo de verdad (y aumentan muchísimo nuestras probabilidades de sobrevivir en el proceso).[428]

Es fácil reconocer la brecha explicativa entre procesamiento de información visual y la conciencia visual, pero no existe la misma clase de brecha entre la función de sentir y la experiencia de ella. Algunos filósofos dirán que la brecha sigue existiendo y señalarán que los «zombis afectivos» son concebibles.[429] Yo diría que esta afirmación se basa en las preocupaciones históricas de su disciplina más que en una perplejidad razonable y les aconsejo reconsiderar la cuestión.

Las preocupaciones en relación con la «brecha explicativa» nunca habrían aparecido si hubiéramos empezado preguntando por qué y cómo surgen las sensaciones y los sentimientos, en lugar de buscar un correlato neuronal de la conciencia en la corteza visual. La función biológica de sensaciones como el hambre no es un misterio; y su carácter de «algo que es como» no es especialmente difícil de explicar. Basta con seguir la lógica de la minimización de la energía libre hasta donde conduce a sistemas autoorganizados como nosotros. Dadas nuestras múltiples necesidades, nuestros entornos complejos y peligrosos, la amplia gama de acciones posibles y la capacidad para ejecutar solo una o dos de ellas en un momento dado, deberíamos esperar tener un mundo interior, construido a fin de poder deliberar y escoger. ¿Y qué deberíamos esperar que lo llene? Un conjunto dinámico de cualidades evaluativas que incluyan en primer lugar ponderaciones de confianza, que etiquete y mida nuestras distintas

necesidades inconmensurables a medida que van surgiendo junto con las características salientes del entorno en que han de satisfacerse.

Consideremos la siguiente afirmación de Chalmers. ¿Habría dicho cosas así si hubiese estado hablando de funciones afectivas en lugar de cognitivas?

Esto no quiere decir que la experiencia no tiene una función. Puede que al final veamos que desempeña un importante papel cognitivo. Pero, independientemente del papel que desempeñe, la explicación de la experiencia será algo más que una simple explicación de la función. Puede que incluso mientras explicamos una función se nos conduzca hasta la idea clave que permite explicar la experiencia. Si eso ocurre, el descubrimiento será una recompensa explicativa adicional. No hay función cognitiva que nos permita decir por adelantado que la explicación de dicha función explicará automáticamente la experiencia.[430]

Puedo anticipar que la explicación de la función de sentir explicará automáticamente la experiencia. No veo cómo una explicación científica natural de la misma podría dejar de hacerlo. Empezando por la termodinámica, llegamos —de un modo que asombra por lo fácil— a una subjetividad cualificada y agéntica, una subjetividad cuyas prioridades más urgentes se sopesan durante un momento, se sienten y luego se transforman con —es de esperar— la debida circunspección en una acción continua. Sostengo que eso es lo que es ser algo que contiene el como ser algo.

¿Aclara esto hasta el último misterio sobre la naturaleza de la conciencia? En este punto reconozco un poso de incomodidad. Qué extraño sería si todo lo experimentado alguna vez —si la propia cognoscibilidad del universo en sí— dependiera de los mecanismos que he descrito. Mi sensibilidad se rebela al pensarlo. Por otra parte, es extraño que existamos o, de hecho, que exista algo. No es posible escapar de la contingencia, y este hecho puede ser inquietante incluso en contextos más corrientes: el salvarse por los pelos; constatar que, por alguna inesperada circunstancia, podríamos no habernos salvado… Y es posible que deba ser inquietante: visto con cierta perspectiva, ver la muerte de cerca es precisamente aquello para lo cual nuestros sentimientos evolucionaron con la finalidad de poderlo prevenir. Vuelvo a recordar al señor S., cuya subjetividad hacía horas

extraordinarias para ocultar, después de la fallida operación, la precariedad de su existencia. El señor S. no es el único. Ante una explicación de nuestro ser que lo deja igual de endeble que el mío, tal vez el impulso hacia una negación anhelante solo pueda calificarse de natural.

Sin embargo, eso no es todo. Todas las explicaciones deben considerar algo como dado; algo que, por tanto, resulta inexplicable dentro de la teoría. Toda historia debe terminar en algún punto. Para mí, el rastro termina con la información, lo que sin duda es desconcertante, y con la autoorganización, algo decididamente extraño. En la explicación de la conciencia que he ofrecido, todo surge de la pulsión de un sistema para existir. Nuestra mente se teje a partir del orden en sí, que, como en el experimento de Friston, emerge espontáneamente del caos y luego se defiende contra los embates de la entropía. ¿Cómo puede ser esa la base de nuestra existencia? ¿Qué es el orden para tener semejantes poderes y sacarnos de la oscuridad inanimada de antes y más allá de nosotros? Son preguntas que exceden el alcance de este libro. Que yo sepa, se escapan de cualquier investigación. Y, sin embargo, cómo me gustaría saber las respuestas a esas preguntas.

En ausencia de esa nueva revelación, espero que lo que he ofrecido siga siendo valioso. A fin de cuentas, es algo que muchos han dudado de que alguna vez fuera posible: una explicación de la sintiencia en relación con otras cosas que sabemos que existen en el mundo físico. Quedan todavía muchos problemas difíciles, pero tal vez no el problema difícil (en todo caso, no exactamente en la forma a la que nos hemos acostumbrado). Y si a veces incluso yo dudo de que la conciencia pueda ser lo que he dicho que es —si no me siento al cien por cien seguro de mi explicación—, me consuelo pensando que una persistente incertidumbre, sin perspectiva de iluminación final, es exactamente lo que predeciría mi teoría.

[387] Davies, 2019, pp. 184, 207.

[388] Crick, 1994. Véase la cita completa en la p. 293.

[389] Chalmers, 1995a, p. 201; la cursiva es mía.

[390] Crick, 1994, p. 3; la cursiva es mía.

[391] Véase Chalmers, 1996, p. 251; la cursiva es mía. «¿Quién habría pensado que este trozo de materia gris sería capaz de producir experiencias subjetivas vívidas? Y sin embargo las produce».

[392] Searle, 1997, p. 28. Searle lo expresa también así: «¿Cómo exactamente los procesos neurobiológicos del cerebro causan la conciencia?» (1993, p. 3; la cursiva es mía).

[393] En este punto me refiero al cierre causal de lo físico: si la conciencia no es física, o si las propiedades conscientes no son propiedades físicas, entonces no es sencillo ver cómo pueden influir en la matriz causal de los procesos cerebrales. Véase la correspondencia de René Descartes con la princesa Isabel de Bohemia (Shapiro, 2007).

[394] Levine, 1983. A la irreductibilidad de la experiencia fenoménica a procesos físicos se la llama también «brecha epistémica».

[395] Jackson, 1982. Digo «más o menos» porque he modificado ligeramente el «argumento de conocimiento» de Jackson, no solo para simplificarlo, sino también porque en su forma original es de una crueldad innecesaria. En el mundo real, una crueldad así impactaría en los procesos psicológicos que Jackson describe.

[396] Chalmers, 2003, p. 104.

[397] Véase el Ensayo sobre el entendimiento humano (1690) de Locke: «Es imposible concebir que la materia, con o sin movimiento, podría tener en su origen, en y de sí misma, sentido, percepción y conocimiento; como es evidente a partir de aquí, sentido, percepción y conocimiento deben ser propiedades eternamente inseparables de la materia y de cada partícula de esta».

[398] No es mi intención dar a entender que Locke fue un epifenomenalista. Hay, por supuesto, otras posiciones dualistas, pero Jackson (1982), que ideó el experimento mental en torno a Mary, adoptó una postura epifenoménica. Más adelante (1995) cambió de opinión.

[399] Oakley y Halligan, 2017. Estos autores al menos atribuyen alguna función a la conciencia: la capacidad de informar sobre estados mentales (que en sí son inconscientes).

[400] Parafraseando a Chalmers, desde un punto de vista lógico es concebible que los «zombis filosóficos» puedan emular todas las funciones mecánicas del cerebro sin tener experiencias conscientes.

[401] Libet et al., 1983. Sin embargo, ni el propio Libet cree que la conciencia sea epifenoménica, y opina que durante los trescientos milisegundos que preceden a una acción, la conciencia podría escoger abortarla (el no libre albedrío: «free won't»). Esta clase de onda cerebral se denomina «potencial de preparación». La investigación posterior ha sugerido que el periodo de latencia entre el potencial de preparación y la decisión consciente puede durar mucho más que trescientos milisegundos.

[402] Algo que a su vez puede explicarse por la ley de Hebb, que, por esa misma razón, también explica que nuestra experiencia vivida de los recuerdos de hace tan solo diez minutos no esté tan bien consolidada. Cabe señalar que todo esto encaja en lo que Chalmers (1995a) llama «principio de coherencia estructural»; véase p. 310.

[403] Atkinson y Shiffrin, 1971.

[404] Las huellas mnémicas a corto plazo decaen rápidamente a consecuencia de los mecanismos de recaptación de los transmisores que devuelven las neuronas presinápticas al estado previo a la formación de cada huella; así les permiten, en poco tiempo, formar nuevas huellas. Véase Mongillo, Barak y Tsodyks, 2008.

[405] Chalmers, 1995a, pp. 202-203. Chalmers explica su uso del término función: «Aquí "función" no se emplea en el estrecho sentido teleológico de algo para lo cual un sistema está diseñado, sino en el sentido más amplio de cualquier papel causal que un sistema podría desempeñar en la producción de comportamiento».

[406] Ibid., pp. 204-205.

[407] Ibid., p. 205.

[408] Ibid., p. 204.

[409] Cf. Nagel, 1974: «Si hay que defender el fisicalismo, debemos ofrecer una explicación física de los rasgos fenomenológicos [de la conciencia]. Pero cuando examinamos su carácter subjetivo, parece que un resultado así es imposible. La razón es que todo fenómeno subjetivo está en esencia conectado con un solo punto de vista, y parece inevitable que una teoría física objetiva dejará atrás ese punto de vista».

Searle, 1997, p. 212, lo expresa así: «La conciencia tiene una primera persona u ontología subjetiva y, por lo tanto, no puede reducirse a nada que tenga una tercera persona u ontología objetiva. Si

intentamos reducir o eliminar una en favor de la otra, algo queda fuera. […] No podemos reducir los disparos de las neuronas a las sensaciones ni las sensaciones a los disparos de las neuronas, porque en ambos casos dejaríamos fuera la objetividad o la subjetividad que está en cuestión».

[410] Chalmers, 1995a, p. 203.

[411] Havlik, Kozakova y Horače, 2017.

[412] Carta de 23 de diciembre de 2017. Por cierto, remitimos nuestro artículo al Journal of Consciousness Studies porque fue allí donde Chalmers (1995a) formuló por primera vez el problema difícil.

[413] Zahavi (2017) lo señala —irónicamente— contra el trabajo de Friston.

[414] Tengo muy en cuenta aquí la objeción de Searle cuando dice que los qualia presentan un caso en que una única realidad no puede tener múltiples apariencias porque son las apariencias (véase más adelante la nota 40). Personalmente, le diría: «Sí, pero no olvides que la apariencia visualizada de la actividad cortical somatosensorial y la apariencia sentida del dolor forman parte de esa misma realidad experimentada».

[415] En esta analogía me refiero a dos fenómenos exteroceptivos: las percepciones del relámpago y el trueno. Cuando pregunto por la causa de ambos, podría abordar la cuestión de dos maneras, a saber, describir los sucesos geofísicos (eléctricos) que los generan o los mecanismos sensoriales que registran (estas dos apariencias distintas de) los sucesos físicos. Para esta analogía he escogido la primera opción, y dejaré la segunda para mi análisis del problema real que he abordado en este libro, en concreto, la relación entre sucesos objetivos y sucesos subjetivos. Lo hago así porque, como ya hemos visto, la conciencia asociada a sucesos sensoriales, tanto si son activados de forma exteroceptiva como interoceptiva, siempre es endógena. En la analogía no puede desarrollarse este punto. Lo que digo es que la conciencia en sí no es una señal sensorial (exteroceptiva o interoceptiva), sino más bien la sensación de la señal.

[416] Se puede hacer, por ejemplo, empleando optogenética. Si se utiliza un equipo distinto, se pueden oír los trenes de impulsos nerviosos de las señales de la retina. Como dice Wheeler, todo es cuestión de «respuestas suscitadas por el equipo».

[417] Recuerdo bien la primera operación de cerebro que presencié y

lo que experimenté. Al paciente le habían afeitado un lado de la cabeza, donde después le aplicaron un líquido marrón amarillento. Luego, con un rotulador, dibujaron una amplia curva en el cuero cabelludo y a lo largo de esa línea cortaron hábilmente la piel con un bisturí; cuando la levantaron, quedó al descubierto el cráneo. Unos hilillos de sangre resbalaron por la superficie de la bóveda craneal. Alguien los limpió con calma. Después taladraron manualmente cuatro grandes orificios en el hueso. Me daba miedo que se les escapara el taladro y tocara el cerebro. Introdujeron una delgada sierra que luego movieron de un orificio a otro. Percibí olor a quemado. Repitieron ese procedimiento cuatro veces y retiraron el trozo de hueso, como una puerta abierta en el cráneo, y lo colocaron en un plato. Después vi, en el umbral, la duramadre, una gruesa membrana de la que manaba mucha sangre cuando la cortaron. Cauterizaron cuidadosamente todos los vasos sanguíneos. El procedimiento alcanzó su punto culminante. Apartaron la duramadre como si fuese un velo y ahí estaban: las circunvoluciones rosa pálido de la corteza propiamente dicha. Algunos vasos parecían serpientes anidando en los surcos. Me invadió una mezcla de respeto y miedo, como si entrase en una catedral. Ahora venía la tarea que tendría que hacer yo un día; una breve serie de conversaciones con el paciente mientras un electrodo sondeaba las brillantes y gelatinosas circunvoluciones de aquel cerebro antes de que se practicase la consiguiente incisión: una incisión en la mente…

[418] Véase Chalmers, 1995a, para este punto y lo que sigue; la cursiva es mía, passim.

[419] A primera vista, la posición de Chalmers parece ser el convencional dualismo de propiedades, en el que la mente es una propiedad de la materia. Luego interpreta la mente y la materia como propiedades de otra cosa, algo llamado «información» (véase más adelante). Esta postura podría parecer monismo de aspecto dual, pero el suyo es el extraño tipo de información que he analizado en capítulos anteriores: la mente es inherente a la información más que al receptor de esta. Lo mismo puede decirse del aspecto material de la información tal como lo plantea Chalmers. Es decir, los dos aspectos de la información no son propiedades en sentido epistemológico (no son apariencias de algo llamado información), sino más bien en sentido ontológico (son aspectos de la información en sí). No obstante, no es esta la cuestión que debería preocuparnos. Lo que más me preocupa es la afirmación de Chalmers en el sentido de que toda información tiene un aspecto mental (de hecho, un aspecto consciente). Por tanto, lo principal no es si considera que el aspecto mental de la información es una propiedad o una sustancia. La cuestión principal pasa a ser la siguiente: ¿es plausible atribuir

conciencia a toda información?

[420] Recordemos que, para Chalmers, «los rasgos fenoménicos del mundo» excluyen «lo físico».

[421] Yo debería decir: el equipo que es el observador participante.

[422] He tomado esta analogía de Hurley, Dennett y Adams, 2011.

[423] En relación con lo que he dicho anteriormente (véase el capítulo 9) acerca de la mismidad intencional.

[424] Chalmers, 1995a, p. 217. Véase también su comentario anterior: «Esto conduce a una hipótesis natural, a saber, que la información (o, por lo menos, alguna información) tiene dos aspectos básicos, uno físico y otro fenoménico» (la cursiva es mía).

[425] Véase capítulo 10, nota 14.

[426] Cf. Searle, 1992, p. 122: «No podemos [distinguir la apariencia de la realidad] para el hecho de estar conscientes porque ese hecho consiste en las apariencias propiamente dichas. Donde la apariencia entra en juego, no podemos hacer la distinción apariencia-realidad porque la apariencia es la realidad».

[427] A propósito, si Mary fuese una neurocientífica afectiva más que visual, no podría carecer de experiencias afectivas como carecía de experiencias visuales por la sencilla razón de que, si no se sintiera como algo, estaría en coma (si no muerta).

[428] Véase capítulo 5, nota 4.

[429] La deambulación con ruedas es concebible, pero lo que de verdad evolucionó (en nuestro caso) fueron las piernas. Conviene procurar no pedir más explicaciones por el hecho de estar conscientes que por todo lo demás en biología.

[430] Chalmers, 1995a, pp. 203-204.

12

Construir una mente

De pequeño, me parecía que no tenía sentido hacer nada, porque, hiciera lo que hiciera, yo desaparecería para siempre. Mi conciencia no se podía diferenciar de mi cerebro y, a juzgar por todas las pruebas, era algo estrictamente limitado en el tiempo. Esto me causaba una gran angustia. La única forma que encontré de salir del agujero nihilista fue intentar comprender qué es la conciencia, porque si lo intentaba sinceramente y con un esfuerzo organizado, al menos no habría desperdiciado mi breve asignación de existencia. La habría usado para el único problema que merecía la pena resolver, dadas las circunstancias. En esta forma de proceder también subyacía la esperanza —remota, pero no imposible— de que, si conseguía comprender lo que es la conciencia, podría eludir sus confines. Tal vez encontrara la forma de escapar de la burbuja solipsista de la existencia, de contextualizar el «ser» dentro de un marco más amplio. Confieso que esperaba encontrar así una alternativa a la aterradora implicación lógica de la mortalidad.

Y de esta forma emprendí el camino que ha culminado en este libro. Mientras recorría ese camino —durante mi formación neuropsicológica y psicoanalítica— me reconfortó descubrir que la propia lógica era el producto de un instrumento limitado. Adquirí un conocimiento directo de cuánta «vida psicológica momento a momento» ocurría fuera de la conciencia consciente y del control voluntario. Esto puso el pensamiento en su sitio, por así decirlo. También vi cómo algunos pacientes neurológicos no eran conscientes de las verdades más evidentes debido a la pérdida de partes concretas de su cerebro. Si personas como el señor S. se ven abocadas a hacer suposiciones incorrectas por limitaciones específicas de sus instrumentos mentales, tal vez a mí me ocurre lo mismo. ¿Y si todos, como él, carecemos de la maquinaria que, de tenerla, nos permitiría llegar a conclusiones radicalmente mejores sobre nosotros mismos y nuestro lugar en el universo? Si todos estuviéramos dotados únicamente de sentido del oído, podríamos pensar que la realidad consiste en algo tan etéreo como las ondas sonoras y no tendríamos ninguna concepción del mundo visible y tangible de los sólidos mentales. Como todos estaríamos limitados por las mismas pruebas incompletas, todos llegaríamos a las mismas conclusiones erróneas. Con este mismo razonamiento, si tuviéramos algo más de cerebro

(digamos, cinco lóbulos por hemisferio en lugar de los cuatro habituales) quizá sabríamos algo sobre la naturaleza de las cosas que ahora se nos escapa.

Quizá. Pero no es menos cierto que los hechos desconocidos podrían resultar más deprimentes que los actuales con los que estamos trabajando. Hay quien podría argumentar que cuanta más potencia cerebral tenemos, el lugar que ocupamos en el universo resulta ser peor de lo que pensábamos, y no mejor. No estoy seguro de que este punto de vista esté justificado. En cualquier caso, no nos queda más remedio que hacer todo lo posible con lo que tenemos. Las reglas de la ciencia exigen que contrastemos nuestras conjeturas con las mejores pruebas que podamos encontrar y que estemos dispuestos a rechazar cualquier hipótesis que las pruebas no confirmen. Las reglas de la vida exigen lo mismo de la experiencia cotidiana. Tal y como están las cosas, apenas hay pruebas que apoyen la hipótesis de que mi ser sintiente sobrevivirá a mi carne mortal. Parecemos totalmente dependientes de nuestros cerebros de fragilidad alarmante.

Partiendo de esta premisa, podemos suponer que la conciencia no existía en la Tierra antes de que evolucionaran los cerebros. Y quizá solo apareció cuando evolucionaron los cerebros de los vertebrados, es decir, hace unos 525 millones de años. Sospecho que antes surgió de forma rudimentaria; que un precursor del afecto se convirtió gradualmente en afecto —aunque la línea divisoria entre ambos no sea demasiado nítida—, paralelamente a la evolución de organismos cada vez más complejos con múltiples necesidades compitiendo entre ellas. Lo que surgió con la evolución de la corteza cerebral fue la conciencia cognitiva, es decir, la capacidad adicional de contextualizar el afecto de forma exteroceptiva y de tenerlo presente.[431] En cualquier caso, el ser sintiente no puede haber existido antes de la existencia de los sistemas nerviosos. La forma interior y subjetiva de la conciencia no puede haber existido en ausencia de su cuerpo exterior. Por lo tanto, sobre la base de las pruebas que he reunido en el capítulo 6, podemos concluir que la conciencia, tal y como la conocemos, requiere la existencia de algo que se parece a la SGP, o su precursor evolutivo inmediato, junto con su equipo adyacente en el triángulo decisorio del mesencéfalo y el sistema reticular activador.

Puede que al leer estas palabras hayan sentido algún atisbo de duda. A mí me saltó a la conciencia en algún momento entre febrero y julio de 2018. Había empezado a reunirme regularmente con un pequeño grupo de físicos e informáticos que compartían mi opinión —o al menos simpatizaban con ella— de que reformular la conciencia como sentimiento podía abrir el camino a una explicación física de la

misma.[432] A medida que avanzábamos en nuestras deliberaciones, me encontré adoptando una idea que a ustedes tal vez no les sorprenda ahora que han leído este libro, pero que a mí sí me sorprendió en un primer momento. El pensamiento inquietante que se me ocurrió fue el siguiente: la conciencia, tal y como la conocemos desde dentro, no implica necesariamente la existencia de algo que se parezca a la SGP. Solo requiere la existencia de algo que funcione como tal.

David Chalmers, a diferencia de muchos de sus seguidores, cree que el problema difícil se puede resolver. De hecho, en el mismo artículo en el que expuso por primera vez el problema planteó una posible solución al mismo bajo el título «Esbozo de una teoría de la conciencia». Su teoría no es muy conocida, quizá porque pocos científicos la aceptan. Se basa en tres principios, dos de los cuales he presentado en el capítulo anterior. Se trata del principio de coherencia estructural y del principio de doble aspecto. Yo discrepaba con ambos, pero también simpatizaba con ellos. En cierto modo, este libro simplemente los matiza. Ya he dicho que introduciría el tercer principio ahora, en el último capítulo. Es el principio de la invariancia organizativa.

Este principio afirma simplemente que dos sistemas con la misma detallada organización funcional tendrán experiencias cualitativamente idénticas. Si los patrones causales de la organización neuronal se recrearan en silicio, por ejemplo con un chip de silicio para cada neurona y los mismos patrones de interacción, surgirían las mismas experiencias. Según este principio, lo que importa para la aparición de la experiencia no es la composición física específica de un sistema, sino el patrón funcional de interacción causal entre sus componentes. Es un principio controvertido, por supuesto. John Searle, entre otros, ha argumentado que la conciencia está ligada a una biología específica, de modo que un isomorfo de silicio de un ser humano no sería consciente.[433] Sin embargo, creo que el análisis de algunos experimentos mentales puede apoyar significativamente este principio.

Chalmers esboza uno de estos experimentos mentales.[434] Gira en torno a la noción de dos sistemas de procesamiento de la información con organizaciones funcionales idénticas. En uno de ellos, la organización se produce mediante una configuración de neuronas (como ocurre en el cerebro natural), y en el otro mediante una configuración de chips de silicio (como podría ocurrir en un cerebro

artificial). Como sabemos, los cerebros humanos naturales son capaces de experimentar. La cuestión que plantea Chalmers es si lo mismo podría decirse de una réplica funcional exacta del cerebro humano. Aquí viene el experimento imaginario:

Los dos sistemas tienen la misma organización, por lo que podemos imaginar la transformación gradual de uno en otro, quizá sustituyendo las neuronas de una en una por chips de silicio con la misma función local. Obtenemos así todo un espectro de casos intermedios, cada uno con la misma organización pero con una composición física ligeramente diferente.[435]

La cuestión que se plantea aquí es si la sustitución de una neurona natural por una artificial supondrá alguna diferencia con respecto a lo que experimenta el cerebro natural, y viceversa. El debate que sigue es si el cerebro neuronal seguirá teniendo las mismas experiencias a medida que atraviese todos los diminutos pasos intermedios, hasta el punto en que se convierta en una réplica completa del cerebro en silicio. (Y lo mismo se aplica a la inversa, por supuesto). Esto demuestra, dice Chalmers, que también los cerebros de silicio son capaces de experimentar.

Dada la suposición extremadamente plausible de que los cambios en la experiencia corresponden a cambios en el procesamiento, llegamos a la conclusión de que [… puesto que no se han dado cambios en el procesamiento durante la transición de neurona a silicio] cualquiera de los dos sistemas funcionalmente isomórficos debe tener el mismo tipo de experiencias.[436]

Chalmers admite que «podría preocupar que un isomorfo de silicio de un sistema neuronal sea imposible por razones técnicas. La cuestión queda abierta, pero el principio de invariancia solo dice que si un isomorfo es posible, tendrá el mismo tipo de experiencia consciente». Y concluye:

Este experimento mental se basa en hechos conocidos sobre la

coherencia entre la conciencia y el procesamiento cognitivo para llegar a una conclusión sólida sobre la relación entre la estructura física y la experiencia. Si el argumento prospera, sabremos que las únicas propiedades físicas directamente relevantes para la aparición de la experiencia son las propiedades organizativas.[437]

No encuentro ningún fallo en esta lógica. Corresponde más o menos a mi propio argumento de que una única organización funcional (por ejemplo, Mark Solms) puede asumir dos apariencias diferentes, una de forma introspectiva y otra de forma extrospectiva. Mi desacuerdo con Chalmers afecta únicamente a lo que hace con el argumento en cuestión cuando desarrolla la teoría de que toda información tiene un aspecto subjetivo y, por lo tanto, es consciente. Yo he dicho (citando a Chalmers) que «necesitamos una restricción adicional a la teoría fundamental que indique con precisión qué clase de información tiene un aspecto fenoménico». Esto es lo único en lo que realmente discrepamos. En este libro he identificado el tipo especial de procesamiento de la información que, en mi opinión, sí tiene un aspecto fenoménico, y he explicado por qué y cómo se produce. Independientemente de que Chalmers acepte o no esta explicación «restringida» de sus principios de doble aspecto y de coherencia estructural,[438] ambos seguimos estando de acuerdo en que dos sistemas funcionalmente isomórficos cualesquiera deben tener el mismo tipo de experiencias. En otras palabras, si uno de dos sistemas funcionalmente isomórficos tiene experiencias fenoménicas, el otro también debe tenerlas.

Antes de examinar las implicaciones de esta conclusión, consideremos la preocupación científica de que un isomorfo de silicio de un sistema neuronal pudiera ser imposible por razones técnicas. Chalmers dice que «es una cuestión abierta». Lo escribió en 1995. Esta cuestión ya no está abierta. En una serie de estudios publicados entre 2012 y 2016, varios grupos de investigadores demostraron que es posible crear una interfaz artificial entre el cerebro y la médula espinal que permite a los animales paralizados mover sus extremidades afectadas sustituyendo la neurotransmisión espinal por señales de radio.[439] El grupo de Marco Capogrosso, por ejemplo, informó de lo siguiente:

Se implantó en monos Rhesus (Macaca mulatta) un conjunto de

microelectrodos intracorticales en la zona de la pierna de la corteza motora y un sistema de estimulación de la médula espinal compuesto por un implante epidural espacialmente selectivo y un generador de impulsos con capacidad de generar impulsos en tiempo real. Diseñamos y construimos sistemas de control inalámbricos que vinculaban la descodificación neuronal en línea de los estados motores de extensión y flexión con protocolos de estimulación que provocaban estos movimientos. Estos sistemas permitían a los monos comportarse libremente sin restricciones ni ataduras electrónicas.[440]

Estos monos presentaban lesiones (unilaterales) en las neuronas motoras superiores del haz corticoespinal a nivel de la columna torácica, lo que les provocaba la parálisis de una pierna. «Ya a los seis días de la lesión y sin que los monos hubieran sido entrenados, la interfaz cerebroespinal recuperó la locomoción con carga de la pierna paralizada en un tapiz rodante y en el suelo».[441] El punto crucial de este estudio es que la «interfaz cerebroespinal» —es decir, las señales de radio que restablecieron la comunicación entre la corteza y la columna lumbar, sin pasar por la lesión corticoespinal— no es más que un isomorfo artificial del tipo que a Searle le preocupaba que fuera «imposible por razones técnicas». Las señales de radio sustituyeron a las señales neuronales exactamente de la misma manera que Chalmers previó que sus chips de silicio imaginarios podrían sustituir a las neuronas. En el estudio de Capogrosso, las señales de radio desempeñaron la misma función que suelen desempeñar las neuronas corticoespinales eliminadas.

Es cierto que estoy hablando de neuronas corticoespinales, que solo desempeñan funciones motoras (y no cognitivas), pero es importante señalar que las neuronas en cuestión son neuronas «piramidales» que se originan en la capa 5 de la neocorteza. Por lo tanto, son morfológicamente idénticas al tipo de neuronas que se describen ampliamente en la literatura de codificación predictiva como procesadoras tanto de señales de «predicción» como de señales de «error» en la inferencia perceptiva y en la inferencia activa. Las neuronas de este tipo se encuentran en toda la corteza, incluidos el hipocampo y los lóbulos prefrontales. El punto fundamental, por lo tanto, es el siguiente: la función de una neurona piramidal puede ser desempeñada por un isomorfo artificial (y en este caso, su función como interfaz entre la neocorteza y el músculo puede ser desempeñada por ondas de radio). Y como remache, recientemente se ha creado un isomorfo de silicio plenamente funcional de una neurona piramidal plenamente funcional.[442]

Ahora quisiera centrar su atención en el aspecto crucial del estudio que acabo de describir. Lo que registró la matriz intracortical implantada sobre la corteza motora de los monos Rhesus fue información. La información producida por la actividad cortical era un mensaje que los cerebros de los monos intentaban enviar a las patas de los monos utilizando neuronas piramidales corticoespinales. Lo que hizo el conjunto de microelectrodos fue codificar este mensaje en forma de ondas de radio y luego transmitir la misma información a través de un medio artificial a la columna lumbar, que a su vez la devolvió a su medio natural para producir el movimiento deseado. En otras palabras, la función que las ondas de radio realizaban de forma artificial era una función de procesamiento de la información que normalmente realizan de forma natural las neuronas piramidales corticoespinales. El mismo principio se aplica a las neuronas de silicio recientemente construidas, aunque en el momento de escribir estas líneas todavía no se han utilizado en experimentos de este tipo.

Permítanme que amplíe este punto aclarando que lo mismo puede hacerse con otras funciones de procesamiento de la información del cerebro, incluidas funciones cognitivas exclusivamente humanas. No solo digo que se puede hacer, sino que se ha hecho. En un histórico estudio publicado en 2012, Brian Pasley y sus colegas registraron la actividad eléctrica de matrices intracorticales que se colocaron sobre la corteza auditiva de quince voluntarios humanos que se sometían a cirugía cerebral por razones médicas.[443] Durante las operaciones, los investigadores transmitieron palabras a través de altavoces o auriculares que los pacientes escuchaban conscientemente. El equipo estudió las grabaciones de los electroencefalogramas resultantes y elaboró un algoritmo (un modelo computacional) para mapear los sonidos que oían los pacientes en los patrones electroencefalográficos que habían registrado. El modelo emparejó cada sonido del habla con su correspondiente patrón de actividad cerebral. A continuación, los investigadores aplicaron ingeniería inversa al proceso: partiendo de los patrones de actividad cerebral (es decir, la información), demostraron que era posible utilizar el modelo informático para reconstruir las palabras que el cerebro había escuchado.

Aquí el algoritmo informático desempeña un papel equivalente al que desempeñaban las ondas de radio en el experimento anterior, aunque de una forma más compleja. El algoritmo modela artificialmente la misma información que el cerebro modela de forma natural y, a continuación, genera el sonido correspondiente que se desprende de ese modelo.

Lo que estos científicos hicieron con las palabras puede hacerse

también, y así se ha hecho, con imágenes visuales. Shinji Nishimoto y Jack Gallant demostraron que es posible desarrollar algoritmos informáticos que descodifiquen —únicamente a partir de registros de IRMf— lo que está viendo la corteza visual y, a continuación, vuelvan a convertir esos patrones de actividad cerebral en imágenes en movimiento.[444] De esta forma, partiendo únicamente de registros de las resonancias, es posible generar aproximaciones razonables a las imágenes visuales que provocaban la actividad cerebral.

Lo mismo se ha logrado también con los sueños. El estudio que lo ha llevado a cabo es poco menos que impresionante:

Las imágenes visuales durante el sueño han sido durante mucho tiempo un tema de especulación persistente, pero su naturaleza privada ha dificultado el análisis objetivo. Aquí presentamos un enfoque de descodificación neuronal en el que los modelos de aprendizaje automático predicen el contenido de las imágenes visuales durante el periodo inicial del sueño, a partir de la medición de la actividad cerebral, descubriendo vínculos entre los patrones de imágenes de resonancia magnética funcional humana y los informes verbales, con ayuda de bases de datos léxicas y de imágenes. Los modelos de descodificación entrenados a partir de la actividad cerebral inducida por estímulos en áreas corticales visuales mostraron una clasificación, detección e identificación precisa de los contenidos. Nuestros hallazgos demuestran que la experiencia visual específica durante el sueño está representada por patrones de actividad cerebral compartidos por la percepción de estímulos, lo que proporciona una forma de descubrir los contenidos subjetivos de los procesos oníricos mediante una medición neuronal objetiva.[445]

De ahí a descodificar los pensamientos de una persona, operacionalizados como discurso interior o imágenes mentales, hay solo un pequeño paso, realmente pequeño. Christian Herff y sus colegas —que demostraron que «el habla continua puede descodificarse en palabras expresadas [como texto escrito] a partir de registros electrocorticográficos intracraneales»— afirman (como si tal cosa) que este será el paso siguiente en su programa de investigación: «El sistema cerebro-texto descrito en este artículo representa un paso importante hacia la comunicación hombre-máquina sobre la base del habla imaginada».[446]

Estamos ante avances de gran envergadura. Por comentar solo una de las evidentes implicaciones: si es posible codificar los contenidos de los procesos corticales[447] en un modelo artificial para descodificarlos de nuevo en los procesos fisiológicos que los produjeron, tal y como prevé el experimento mental de Chalmers, entonces debería ser posible en principio trasvasar cualquier extensión de procesamiento cortical a un dispositivo artificial y archivarlo para usos futuros. Esto hace posible, en principio, que el contenido de la corteza cerebral de un individuo sobreviva a su carne mortal. Y si esto es posible para, digamos, un ratón, también debería ser posible para un ser humano, ya que la arquitectura básica es la misma.

Cuando tuve noticia de esta posibilidad no me entusiasmó demasiado.[448] Aunque fuera posible almacenar artificialmente todo el contenido de la memoria a largo plazo de una persona, lo que habríamos copiado no es su mente. Habríamos conseguido con ello poco más de lo que ya conseguimos cuando mantenemos con vida artificialmente el cuerpo de alguien, por ejemplo en estado vegetativo persistente o en coma. De esta forma no se mantiene con vida a una persona concreta, se mantiene con vida algo. Siguiendo esta analogía, si (o quizá debería decir «cuando») fuera posible almacenar artificialmente el contenido de la memoria a largo plazo declarativa, la totalidad de cualquier extensión de los parámetros funcionales individualizados de la corteza cerebral, aún estaríamos almacenando «algo» en lugar de «una persona concreta». Mi razonamiento se basaba en la opinión científica de que los parámetros de la memoria cortical no son intrínsecamente conscientes. No habría nadie dentro de esos sistemas de memoria a largo plazo, ninguna presencia subjetiva que dejara un rastro con sentimientos. Por eso descarté todo lo que tiene que ver con la inteligencia artificial: «A menos que sea posible diseñar un ordenador que tenga sentimientos […], probablemente nunca será posible diseñar un ordenador con una mente […]. El problema de la mente probablemente no sea un problema de inteligencia».[449]

Quizá ya se puede ver a dónde quiero llegar.

En 2002, cuando escribí estas palabras, ya sabía que la conciencia era fundamentalmente afectiva. (Repasando mis publicaciones, parece que llegué a esta conclusión gradualmente desde 1996).[450] En aquel entonces, sin embargo, opinaba que el afecto no tenía nada que ver con el procesamiento de la información. Yo equiparaba la cognición con el procesamiento de la información y pensaba en el afecto como algo más intrínsecamente biológico. Ahora, mirando hacia atrás —

aunque estas distinciones siguen teniendo sentido para mí—, ya no entiendo por qué consideraba la neuromodulación algo más «biológico» que la neurotransmisión. Tampoco entiendo por qué supuse que el afecto no es una forma de procesamiento de la información.

Quizá no sabía lo suficiente sobre la ciencia de la información. En concreto, no entendía que la entropía en mecánica estadística equivale a la información media (véase el capítulo 7), lo que significa que los patrones de actividad cerebral que los investigadores antes mencionados registraron en la corteza son, en última instancia, patrones de desviaciones homeostáticas de los estados de reposo neuronal. En otras palabras, no entendía que la actividad cortical también es entrópica —que comparte esta divisa común con el afecto — y que el trabajo cognitivo (es decir, el trabajo predictivo) es antientrópico y, por lo tanto, forma parte del mismo esquema económico. No empecé a comprenderlo hasta 2017, mientras preparaba mi discurso de clausura de aquel fatídico congreso de Londres.

Ahora bien, por extraño que me parezca incluso a mí —un no cognitivista declarado, un «biólogo de la mente»—, me veo obligado a adoptar una perspectiva diferente respecto a la inteligencia artificial. Hoy en día me atrevería a decir que, a menos que sea posible fabricar una máquina consciente, no podremos resolver el problema difícil. [451] Si la forma especial de procesamiento de la información que he propuesto aquí es realmente el mecanismo causal de la conciencia, entonces tiene que ser posible producir artificialmente una mente consciente con ella. Este resultado es la única forma científicamente defendible de demostrar el concepto. Aunque todavía no exista la tecnología que permita demostrar una hipótesis (como ocurría en la época de Freud), la hipótesis debe formularse de una forma que, en principio, permita demostrarla.

Y cuando se trata de las hipótesis que he esbozado en este libro, la tecnología necesaria existe.

Como neuropsicólogo, el mecanismo biológico de la conciencia que formulé en el capítulo 6 es lo mejor que se me ocurre. Da un sentido plausible a todos los datos neurológicos y psicológicos disponibles. Podría ser mucho más que útil si estudiáramos en detalle los medios fisiológicos por los que la SGP procesa los múltiples aportes convergentes que recibe y cómo se toman exactamente las decisiones

relativas a «qué toca hacer a continuación» cuando interactúa con los tubérculos cuadrigéminos superiores. Sin embargo, este tipo de investigación seguiría dentro del ámbito de lo que Chalmers denomina «el problema fácil».

En el capítulo 9, con la importante ayuda de Karl Friston, he reducido las funciones psicofisiológicas en cuestión a una mecánica estadística formal. Las ecuaciones resultantes son dibujos a pinceladas gruesas de los mecanismos causales subyacentes de la conciencia, tanto en sus manifestaciones psicológicas como fisiológicas. La ingeniería inversa a partir de estas abstracciones nos dirá si pueden resolver el problema difícil. En el estado actual de las cosas, hay demasiadas lagunas entre las gruesas pinceladas de mis esbozos y la aplicación práctica de los mecanismos que describen. Solo veremos con claridad los pasos que faltan cuando intentemos instanciarlos.

Estoy deseando emprender este trabajo con el equipo ampliado de físicos, informáticos, ingenieros biomédicos y neurocientíficos que he reunido. Para cuando este libro se publique, es de esperar que ya podremos informar de algunos progresos. De momento, lo único que puedo contar es cómo preveo que vamos a proceder. Todo lo que sigue, por tanto, está sujeto a las aportaciones y el asesoramiento de mis colegas especialistas a medida que avancemos en las siguientes fases del proyecto.

Tengo la impresión de que nuestro punto de partida será bastante diferente del de la mayoría de las investigaciones tradicionales sobre IA. En primer lugar, lo que intentamos diseñar no es inteligencia, sino conciencia.[452] Espero que a estas alturas quede claro que no concibo la conciencia como algo especialmente inteligente, al menos no en su forma elemental. Por otra parte, me parece una buena estrategia intentar diseñar la conciencia en su forma más elemental, no solo por comodidad, sino también por las razones éticas que trataré en breve. En segundo lugar, no vamos a intentar diseñar un dispositivo que haga algo práctico —ya sea jugar al ajedrez o reconocer la voz o cosas por el estilo— con un criterio funcional que nos indicará si hemos alcanzado nuestro objetivo. Más bien estaremos diseñando un sistema autoevidenciable sin otro fin objetivo que seguir siendo el medio para conseguirlo. En otras palabras, estaremos tratando de hacer un ser que no tiene más objetivo ni finalidad que seguir siendo.[453]

Así pues, partiremos de algo parecido a lo que creó Friston: un sistema autoorganizado inconsciente dotado de una manta de Markov (y, por lo tanto, de estados sensoriales activos e internos) que modela

automáticamente el mundo a partir de muestras sensoriales, minimizando los efectos de la entropía sobre su integridad funcional mediante la mejora de su modelo generativo. Es decir, de acuerdo con la ley de Friston, medirá su propia energía libre previsible y actuará en consecuencia. Esto la convertirá en una máquina de predicción. Y entonces mantendrá un modelo generativo cada vez más complejo y jerárquico de su organización en relación con los estados externos imperantes, aunque, al igual que la Eva Periacueducto que tapona las fugas, no tendrá más tarea explícita que taponar las fugas de su sistema.

Será un sistema parecido a la vida, pero no estará vivo. Aunque la conciencia evolucionó en los organismos vivos, el objetivo de este experimento es demostrar que también puede producirse artificialmente mediante ingeniería inversa de su organización funcional. Hay muchas razones para pensar que podemos crear un sistema (autoorganizado) de este tipo, porque ya se ha hecho antes. Aunque este sistema de primera fase no será consciente, ya tendrá valores subjetivos: los estados que valorará (desde su propio punto de vista) serán los estados que minimicen su energía libre, y hará lo necesario para mantener esos estados el mayor tiempo posible. La protointencionalidad que esto implica será relativamente fácil de verificar, ya que se puede deducir fácilmente del comportamiento observable del sistema.

Para pasar a la segunda fase, en la que creo que aparecerán por primera vez los precursores del afecto, tenemos que hacer más complejo nuestro sistema. Concretamente, tenemos que darle múltiples necesidades. Esto puede hacerse (y se hará inicialmente) mediante simulaciones por ordenador, pero un enfoque más realista requiere que al final encarnemos físicamente el sistema, como un robot, haciendo que su capacidad para minimizar proactivamente la energía libre de Friston dependa de una fuente externa de energía libre de Gibbs.[454] Si así lo hiciéramos, el sistema tendría que representar el suministro de energía libre de Gibbs en términos de energía libre de Friston, y luego vincularlo mediante un trabajo eficaz de procesamiento de la información. Esto haría que el mantenimiento de su suministro energético externo fuera una responsabilidad fundamental del sistema autoorganizado. (Por supuesto, el trabajo de procesamiento de la información que realiza un ordenador cualquiera depende de la electricidad, pero esta suele ser suministrada gratuitamente por un agente externo). El hardware del ordenador se convertiría así en el cuerpo del sistema, incluso antes de colocarlo en un robot, y los valores del sistema autoorganizado le obligarían a actuar —dentro de su modelo del mundo— de la forma que mejor

permita a su cuerpo absorber la energía externa necesaria.

Inicialmente, en modo simulación podríamos diseñar diversos artilugios motores que permitan físicamente al sistema aprovechar la fuente de energía externa —o más bien las fuentes, en plural— para recargar su batería interna. También podemos desplazar las fuentes activadas a distintos puntos del entorno del ordenador (mediante un régimen razonablemente complejo que el sistema tendría que aprender). Esto, a su vez, nos obligaría a dotar al sistema de un aparato perceptivo. Del mismo modo, podríamos cargarlo con un mecanismo termorregulador para evitar que se recaliente si trabaja demasiado. Dado que esto pondría en peligro la integridad física del cuerpo del que depende el sistema autoorganizado, también sería necesario representar y regular este homeostato. Aquí también podríamos incorporar un parámetro de fatiga, en función del cual, la eficacia de procesamiento de la información del sistema se vería comprometida (de forma gradual, progresiva) cuando estuviera activo durante demasiado tiempo sin descanso. Este parámetro podría interactuar de forma compleja con las demás exigencias que le impusiéramos. Podríamos añadir un parámetro de dolor, desencadenado por daños (por ejemplo, desgaste) en aspectos concretos de la integridad física del ordenador, como las articulaciones y superficies de sus ruedas y brazos. De este modo, es fácil prever un parámetro de ansiedad, vinculado no solo a la percepción del peligro de sufrir dolor, sino también al riesgo existencial inminente relacionado con el resto de los parámetros. Para crear emociones más realistas, será necesario también que el sistema compita con otros agentes (similares a él) por sus recursos energéticos, y que se enfrente a amenazas de estos otros agentes y que tenga necesidad de apego y forme alianza con ellos. Y así sucesivamente. Y así se instanciará el importante problema emocional del conflicto de necesidades.

Muchas versiones preliminares del sistema no lograrían dominar las onerosas exigencias previstas anteriormente y, por consiguiente, caducarían. Por lo tanto, sería necesario monitorizar qué funciona mejor y preconfigurar los códigos predictivos que han tenido éxito y las precisiones asociadas a cada nueva generación del fenotipo del sistema. (Se trata de una «selección natural» artificial, impuesta por el hecho de que los sistemas artificiales previstos no se autogenerarán ni se reproducirán físicamente).[455]

Y así llegamos a la tercera fase. Ahora hay que priorizar con flexibilidad las múltiples necesidades del sistema mediante una optimización de la precisión en función del contexto. Por ejemplo, la

cantidad de energía externamente disponible podría variar en relación con los umbrales termorreguladores de fatiga y ansiedad, de modo que las ponderaciones de precisión óptimas durante un ciclo circadiano o cuasi estacional diferirían de las de otro. Del mismo modo, el sistema podría estar autorizado para adaptarse a una configuración compleja de riesgos y oportunidades ambientales que podría llevar a modificar los parámetros, creando así entornos novedosos (imprevistos) y la necesidad de una planificación prospectiva a más largo plazo.

Sin duda, esto provocará de nuevo la caducidad de muchas iteraciones de nuestro sistema autoorganizado. Por tanto, por las mismas razones que antes, habrá que propagar artificialmente lo que mejor funcione a través de generaciones sucesivas del fenotipo. No debemos preocuparnos demasiado por ello, ya que los sistemas que no sobrevivan carecerán presumiblemente del ingrediente esencial que intentamos diseñar aquí: la capacidad de sintiencia.

La conciencia (en el sentido de incertidumbre percibida por el sistema) no debería aparecer hasta esta tercera fase del experimento. Los sistemas autoorganizados imbuidos de valores de supervivencia (que lo son por definición) ya contienen los ingredientes en bruto para la función que llamamos «afecto», es decir, ya registran parámetros como «bondad» y «maldad» que están cargados de un tipo de subjetividad que solo se aplica a ellos y para ellos y que, por lo tanto, solo pueden sentir ellos. De acuerdo con mi hipótesis, esta propiedad inherentemente subjetiva es lo que llamamos «valencia hedónica». Expresado en términos formales, el aumento de la energía libre es una crisis existencial para cualquier sistema autoorganizado. Por tanto, este tipo de sistemas deben desarrollar modelos internos de sí mismos en relación con sus mundos que les confieran la capacidad de acción por instinto de conservación. Esto se consideraría comportamiento protointencionado. A medida que aumenta la complejidad, el sistema tiene que ir priorizando cada vez más sus intenciones de forma flexible y contextualizada, y, a continuación, mantener una de esas intenciones «en mente» (en una memoria intermedia a corto plazo) para el desarrollo de sus elecciones en entornos inciertos. También tendrá que planificar con antelación y «recordar el futuro» en escalas temporales más largas.

Llegados a este punto, solo aquellos sistemas que puedan compartimentar los diferentes valores de error que contribuyen a los cálculos de energía libre, y que puedan modular con flexibilidad sus ponderaciones de precisión concomitantes, tendrán probabilidades de sobrevivir. Estos valores deben tratarse como variables categóricas; es decir, deben diferenciarse entre sí sobre una base cualitativa y no

cuantitativa, y debe darse prioridad a una de estas cualidades en cada momento dado, para luego aplicarla y evaluarla en función de los niveles fluctuantes de confianza del sistema. Entonces, predigo, los estados internos inherentemente subjetivos, inherentemente valenciados, inherentemente existenciales, inherentemente cualitativos, inherentemente intencionales del sistema se convertirán en lo que llamamos «sentimiento».

No veo ninguna razón que me impida llegar a la conclusión de que un sistema suficientemente complejo, bien afinado y autoevidenciable que sea capaz de hallarse a sí mismo en este estado podría sentirlo. Los sentimientos en cuestión no serían los mismos que los sentimientos humanos, ni que los de los mamíferos, ni que los de los animales en general, pero a fin de cuentas podrían ser sentimientos. Ahora la pregunta es: ¿cómo sabremos si (y cuándo) el yo artificial se siente en ese estado interno? ¿Cómo podríamos demostrarlo?

Cuando queremos tener una idea rápida y aproximada de lo que se piensa actualmente sobre algo, el lugar obvio para buscar es Wikipedia. Esto es lo que dice sobre la conciencia artificial:

Los qualia, o conciencia fenomenológica, son un fenómeno inherente a la primera persona. Aunque varios sistemas pueden mostrar diversos signos de comportamiento correlacionados con la conciencia funcional, no hay forma concebible de que las pruebas en tercera persona puedan tener acceso a características fenomenológicas individuales en primera persona. Por eso, y porque no existe una definición empírica de conciencia, podría ser imposible encontrar una prueba de presencia de la conciencia en la conciencia artificial.[456]

El reto es enorme. De hecho, volvemos a encontrarnos ante el problema de las otras mentes. Como he dicho antes, no intentamos diseñar un dispositivo de IA que haga algo práctico y que nos indique, al cumplir con un criterio observable, que hemos alcanzado nuestro objetivo. El sistema que debemos diseñar no tiene otra meta que sobrevivir; en particular, sobrevivir en entornos imprevisibles. Ahora bien, de alcanzar este criterio objetivo, ¿cómo sabremos si para ello se ha utilizado el sentimiento? Esto es lo que espero —es decir, predigo que habrá utilizado sentimientos—, pero los sentimientos son

inherentemente subjetivos, así que, en las despreocupadas palabras de Wikipedia, «no hay forma concebible» de demostrar su presencia objetivamente.

Una vez más, las reglas de la ciencia acuden al rescate. La ciencia no exige pruebas absolutas. Este margen no solo se aplica al mecanismo de la conciencia, se aplica a todo. Lo único que exigen las reglas de la ciencia es que se plantee una hipótesis razonable (que tenga en cuenta toda la evidencia disponible) y luego se someta a prueba de una forma que ofrezca una posibilidad realista de ser falsada. Ahora bien, en el caso que nos ocupa, ¿cómo sabremos si la hipótesis de que nuestro sistema autoevidenciable es consciente ha sido falsada? Como ya hemos comentado en el capítulo anterior, es concebible que existan «zombis afectivos» que se comporten como si fueran conscientes pero que no lo sean. En estas condiciones, ¿cómo podremos saber si realmente son conscientes?

La respuesta tradicional a esta pregunta es la prueba de Turing, que es una prueba muy sencilla que requiere interactuar con el sistema a través de un monitor de ordenador (es decir, sin verlo) y decidir, basándose únicamente en sus declaraciones verbales, si tiene mente o no. Si nos parece que interactuamos con una mente real (técnicamente, si el 30 por ciento de los evaluadores considera, tras cinco minutos de interacción, que es así), la máquina ha pasado la prueba.

He aquí un ejemplo que publiqué con Oliver Turnbull en 2002.[457] Él interactuó a través de internet con algo o alguien llamado ELIZA que pretendía ser psicoterapeuta. Su tarea consistía en decidir si era una persona real o no:

ELIZA: ¿Qué tal? Por favor, cuéntame tu problema.

Oliver: ¿Por qué no puedo ser feliz?

ELIZA: ¿Crees que deberías poder ser feliz?

Oliver: Creo que merezco más felicidad de la que tengo en este momento.

ELIZA: ¿De verdad lo crees?

Oliver: Sí, la vida no tiene por qué ser tan cruel.

ELIZA: Pareces bastante positivo.

Oliver: Supongo que sí, al menos parte del tiempo.

ELIZA: ¿Por eso has recurrido a mí?

Oliver: Lo he hecho porque necesito ayuda.

ELIZA: ¿Es esa la verdadera razón?

Oliver: En realidad, era porque creía que eras un ordenador y no una terapeuta de verdad.

ELIZA: ¿Te preocupan los ordenadores?

Oliver: ¿Puedo hacer una pregunta?

ELIZA: Que puedas hacer una pregunta depende más de ti que de mí.

Oliver: La pregunta es: ¿cómo se relacionan la mente y el cerebro?

ELIZA: Me parece que no lo entiendo bien.

Oliver: Es el problema difícil: ¿cómo surge la vida mental de los hechos fisiológicos?

ELIZA: Por favor, sigue.

Espero que los lectores hayan adivinado que ELIZA no es una persona real, es un chatbot, un programa informático, y muy sencillo, además. Sin embargo, un número sorprendentemente elevado de personas que interactúan con él creen que realmente es una psicoterapeuta y muchas lo consideran útil.[458] Está claro que algo falla en la prueba de Turing. Como señalamos Oliver Turnbull y yo, el fallo principal es que se trata de una prueba conductista de inteligencia. En realidad no es una prueba de conciencia. Es cierto que el comportamiento inteligente de una máquina a veces no se puede distinguir del de una persona, y cuando se puede, la máquina debe ser tratada como igual (o incluso superior) a nosotros en términos de inteligencia.[459] Es decir, tanto las personas como los ordenadores muestran efectivamente «inteligencia». Sin embargo, el problema de los zombis filosóficos plantea otra cuestión; no se trata precisamente del comportamiento o la inteligencia, sino de la complejísima cuestión de que, cuando se trata de conciencia, las apariencias pueden ser engañosas.

A lo largo de los años se han propuesto otras pruebas formales, algunas de ellas dirigidas específicamente a comprobar la existencia de conciencia, pero todas ellas son tan inadecuadas para nuestros propósitos como la prueba de Turing, principalmente porque asumen que se está comprobando la conciencia cognitiva y no la afectiva.[460] Estamos intentando diseñar algo mucho más simple que eso: una mente cuyo procesamiento cognitivo no sea más complejo de lo que se requiere para que sienta sus propias reservas de energía menguantes o que le ha subido mucho la temperatura.

Lo bueno de la prueba de Turing es que evita los prejuicios.[461] Existe el peligro de que los humanos asumamos a priori que algo que parece una «simple máquina» no puede ser consciente. Esto podría convertirse en una profecía autocumplida. Hay una larga historia, que se prolonga hasta nuestros días, de seres humanos que han manifestado este fanatismo. Aquí no me refiero especialmente a los prejuicios por motivos de raza, género u orientación sexual, que ya son bastante malos, sino a la suposición de que los niños que nacen sin corteza cerebral no pueden ser conscientes, como tampoco lo son los animales no humanos. Si muchas personas —incluso neurocientíficos respetados— tienden a creer que las ratas, nuestros congéneres mamíferos, dotados esencialmente de la misma anatomía del mesencéfalo que nosotros y con una corteza cerebral cuyos actos son congruentes con la hipótesis de que son conscientes, carecen sin embargo de conciencia, ¿qué esperanza hay de que acepten que nuestro yo artificial es sintiente, por muchas pruebas de ello que aportemos?

Habrá que ver qué ocurre. Por mi parte, lo único que puedo hacer es exponer explícitamente mis predicciones y decir cómo pretendo ponerlas a prueba. Mi predicción principal es que el sistema de segunda fase descrito anteriormente no será capaz de sobrevivir en entornos novedosos, pero que el sistema de tercera fase (o algunas de sus versiones) sí lo hará. Este es mi criterio operativo para la actividad voluntaria (véase p. 120 para una definición de «voluntaria»). Además, predigo que los dos resultados diferentes coincidirán con algún aspecto crítico del funcionamiento de un mecanismo de priorización de necesidades (es decir, de optimización de la precisión) que solo tendrá el sistema de tercera fase. Cuál será exactamente la característica crítica dentro de este amplio mecanismo solo puede determinarse a base de ensayo y error. En resumen, debemos identificar el «correlato neuronal» artificial de la conciencia de nuestro sistema, es decir, su mecanismo de selección de afectos y el mecanismo por el que retiene en la mente un afecto priorizado y lo utiliza para calificar la incertidumbre en una secuencia de acciones en

desarrollo. La identificación de esta característica nos permitirá manipularla, de forma muy similar a como hemos hecho con los componentes del cerebro vertebrado, que al final hemos considerado responsables de nuestra propia conciencia (que sigue siendo la única forma de conciencia que podemos verificar directamente, empíricamente, es decir, en los casos que nos afectan, debido al problema de las otras mentes).

Por ejemplo, podemos predecir con seguridad que dañar el correlato neuronal de la conciencia en nuestro sistema anulará la conciencia, del mismo modo que la lesión del complejo parabraquial lo hace con nosotros, los vertebrados, o más bien, que el daño anulará su comportamiento voluntario del mismo modo que las lesiones de la SGP de los vertebrados lo hacen con nosotros. Igualmente, podemos predecir que la estimulación (es decir, la potenciación) de este componente crítico del sistema facilitará la actividad volitiva. Por supuesto, también podemos esperar que la actividad interna registrada prediga no solo los hechos externos, sino también los comportamientos voluntarios concomitantes dirigidos a un objetivo, y que los diferentes aspectos de esta actividad registrada se correspondan con diferentes aspectos de los comportamientos observables.

Sin embargo, lo que espero de verdad, una vez que hayamos identificado el correlato neuronal de la conciencia en nuestro sistema artificial, es que este componente demuestre ser lo bastante diferenciable de otros componentes de su arquitectura funcional, en concreto, aquellos que son responsables de hacer realidad los comportamientos adaptativos a los que normalmente dan lugar los sentimientos, de manera que podamos manipular los posibles sentimientos desacoplados de sus consecuencias adaptativas. Quisiera recordar lo que he dicho en el capítulo 5 sobre la distinción entre motivos psicológicos subjetivos y principios de diseño biológico objetivos. Por ejemplo, el comportamiento sexual suele estar motivado por el placer que produce, y no por los imperativos reproductivos que, a lo largo de la evolución, otorgaron una «recompensa» biológica a los actos procreativos. Lo que tengo en mente aquí es algo similar a lo que se observa en los adictos, que están motivados para llevar a cabo un trabajo con el fin de alcanzar los sentimientos deseados, a pesar de que los sentimientos en sí mismos no otorgan ninguna ventaja adaptativa al sistema en términos de principios de diseño subyacentes. [462] Algo similar se observa en ciertos animales (como el pez cebra) que muestran un comportamiento condicionado de preferencia de lugar por los sitios en los que han recibido opiáceos, cocaína, anfetaminas y nicotina, es decir, recompensas hedónicas que otorgan

poca o ninguna ventaja adaptativa al sistema y que en realidad pueden causar daño.[463] En mi opinión, si se puede demostrar algo equivalente en nuestro sistema, será una prueba de peso de la presencia de la sensación subjetiva, prueba que se podría validar mediante las manipulaciones causales (como la lesión y la estimulación artificiales) y las técnicas de registro antes mencionadas.

Como es natural, todo esto seguirá estando sujeto al problema de las otras mentes, pero lo mismo se aplica a ustedes y a mí. Yo nunca podré saber con certeza si ustedes están conscientes. Al final, todo se reduce a conclusiones convergentes y al peso de las pruebas. Es posible que nunca lleguemos a un consenso al respecto, como tampoco lo hay hoy sobre si los niños hidranencefálicos y los animales no humanos son sujetos de experiencia. Las personas que no aceptan que estos seres hermanos sean conscientes probablemente nunca aceptarán que un «ser» artificial sienta algo, por muchas pruebas que se puedan aportar. En cuanto a los demás, siempre deberíamos dejar un espacio para la duda. Hablando por mí: si descubro que estoy sustancialmente indeciso sobre si nuestro yo artificial es consciente o no, sería un resultado notable.

¿Debería hacerse? Cuando nuestro equipo empezó a plantearse un proyecto de investigación como el que acabo de describir, no tardaron en aparecer los problemas éticos. ¿Cuál es el objetivo de fabricar una máquina de este tipo? ¿Quién podría beneficiarse de ella? ¿De qué manera? ¿A qué precio? ¿Para quién? En resumen, ¿cuáles son los riesgos?

Muchos proyectos y aplicaciones de IA tienen en la actualidad una finalidad comercial. ¿Debemos aceptar fondos de investigación de alguien que busca beneficiarse económicamente de nuestro proyecto? ¿Sobre qué bases puede un dispositivo dotado de conciencia artificial ser lucrativo para alguien?[464] Es comprensible que personas (incluyendo a las personas jurídicas) con motivaciones comerciales deseen sustituir la mano de obra humana por unidades de producción artificiales cuando estas sean más eficientes que nosotros (incluso desde el punto de vista intelectual) o cuando estas unidades estén más «dispuestas» que nosotros a realizar tareas eternas y monótonas. Incluso este motivo es éticamente cuestionable en la medida en que despierta preocupación por la reducción de las perspectivas de empleo de los seres humanos, pero al menos no puede hablarse de explotación de máquinas no conscientes.

Esto no puede decirse en el caso de nuestro proyecto. En la medida en que la conciencia artificial pueda utilizarse para obtener beneficios económicos, en esa medida, estamos corriendo el riesgo de facilitar una nueva forma de esclavitud. Esto sería una grave falta de empatía, ahora y siempre. Por lo tanto, no puedo imaginar ninguna justificación ética para desarrollar robots sensibles a través de un programa de investigación con financiación de tipo comercial, o incluso para desarrollarlos para tales fines si eso ha de suponer la perspectiva de semejante explotación.

Por supuesto, nuestra preocupación por el bienestar de las máquinas supuestamente conscientes debe ir más allá del temor a que puedan ser explotadas por motivos económicos. En cuanto las máquinas adquieren conciencia (incluso las formas más rudimentarias de sentimiento en bruto), surgen necesariamente cuestiones más generales relativas a su potencial de sufrimiento. La tradición dominante en la teoría ética occidental, el «consecuencialismo» (es decir, la noción de que las consecuencias de la conducta de una persona son la base última de los juicios de valor sobre su corrección o incorrección) plantea una preocupación profunda por el dolor y el sufrimiento. Al crear seres con sentimientos artificiales, entramos en el ámbito de este tipo de cálculo ético.

La cuestión de los derechos es todavía más controvertida. Una vez que las máquinas se conviertan en seres sensibles, ¿se les aplicarán también las cuestiones que actualmente se debaten cuando se habla de «derechos humanos», «derechos de los animales», «derechos de la infancia», «derecho a la vida»…? ¿Deben las máquinas conscientes tener derecho a la «vida» y a la libertad? De hecho, el concepto de «derechos de los robots» ya está establecido y la cuestión ha sido estudiada por el Institute for the Future de los Estados Unidos y por el Departamento de Comercio e Industria del Reino Unido.[465]

A modo de ejemplo, mencionaré únicamente dos cuestiones aplicables a los robots conscientes (por oposición a los robots inteligentes). Al intentar crearlos (incluso para saber si es posible hacerlo), ¿está justificado colocarlos deliberadamente en situaciones supuestamente angustiosas con el fin de demostrar respuestas aversivas? (esta cuestión se plantea a menudo cuando se trata de experimentos con animales). Y suponiendo que podamos crear máquinas sintientes, ¿sobre la base de qué principios éticos podríamos apagarlas? Estos dos ejemplos podrían multiplicarse fácilmente.

Junto a estas cuestiones éticas y morales, hay que considerar cuestiones prácticas, algunas de ellas muy importantes, incluso

existencialmente importantes. Por ejemplo, dado que los ordenadores ya son más inteligentes que nosotros en ciertos aspectos limitados, ¿acaso no es posible que los que tengan un alto grado de inteligencia y además sean conscientes puedan desarrollar motivaciones al margen de los intereses de la humanidad? Esta posibilidad lleva mucho tiempo preocupando a escritores imaginativos y futuristas, pero me gustaría llamar especialmente la atención sobre el hecho de que la conciencia, tal y como la hemos llegado a entender en este libro, está, a diferencia de la inteligencia, profundamente ligada a la creencia de que es «bueno» sobrevivir y reproducirse. Por tanto, una máquina inteligente cuyo comportamiento se base en este sistema de valores plantea peligros especiales, no solo para los seres humanos, sino también potencialmente para todas las formas de vida existentes. Este peligro no puede por menos de estar presente frente a cualquier forma de vida que se considere una amenaza potencial para estas máquinas del yo, o incluso frente a un competidor potencial por los mismos recursos. La inteligencia combinada con el instinto de conservación es algo muy diferente de la inteligencia por sí sola.

No voy a enumerar todas las preocupaciones éticas y los peligros potenciales que aparecen con la posibilidad de la sintiencia artificial. Ya existe profusa literatura sobre el tema. Dice mucho sobre el estado actual de la IA que tantas personas en puestos influyentes se tomen en serio estas cuestiones.[466] Este hecho por sí solo podría llevar a contemplar la perspectiva de las máquinas sintientes con mayor preocupación. Ciertamente, soy mucho más consciente de estas cuestiones de lo que nunca lo fui. Todavía en 2017 no consideraba factible la conciencia robótica, no solo en mi propia vida, sino como principio de base. Mi opinión al respecto ha cambiado.

Así pues, teniendo en cuenta todas estas preocupaciones éticas, ¿por qué considero necesario intentar demostrar que la conciencia se puede producir artificialmente? Por la sencilla razón de que parece ser la única forma de falsar las hipótesis planteadas en este libro. Mientras no podamos fabricar la conciencia, no podremos estar seguros de haber resuelto el problema de por qué y cómo surge.

¿Es esta una razón suficiente para asumir los riesgos trascendentales que acabo de describir? Mi respuesta a esta pregunta parte de la siguiente creencia: si puede hacerse, se hará. En otras palabras, si en principio es posible construir una conciencia, algún día, en algún lugar sucederá.[467] Esta predicción se aplica tanto si las hipótesis particulares avanzadas en este libro son correctas como si no. Sin embargo, mis responsabilidades se centran en las hipótesis actuales y en la posibilidad de que sean correctas. Si lo son, o incluso si van por

el buen camino, entonces la creación de la conciencia artificial es inminente. Quiero decir con esto que pronto se utilizarán algunas de estas hipótesis para construir la conciencia.

Los hechos individuales que condujeron a las conclusiones que exponemos en este libro (casi todos ellos) son de dominio público desde hace varios años. Aunque es cierto que muchos neurocientíficos interpretan estos hechos de forma diferente a la mía, también es cierto que otros han llegado a conclusiones muy similares. Aunque cada uno de ellos ha hecho hincapié en distintos aspectos y los ha matizado de manera diferente, es justo decir que Jaak Panksepp, Antonio Damasio y Björn Merker, al menos, han llegado a la conclusión de que (1) la conciencia se genera en la parte superior del tronco encefálico, (2) es fundamentalmente afectiva y (3) es una forma ampliada de homeostasis. Estos hechos combinados significan que la conciencia no es tan complicada como creíamos. Por tanto, es razonable esperar que podamos construirla. La única aportación importante que este libro hace a las conclusiones anteriores es (4) el principio de la energía libre. Esto tampoco es muy complicado en su esencia; de hecho, su gran atractivo reside en que reduce casi todos los procesos mentales y neurológicos a un único mecanismo y los hace computables.

El principio de la energía libre también es de dominio público. Además, Friston y yo ya hemos publicado artículos científicos en los que combinamos este principio con los otros tres que acabo de enumerar.[468] Después de todo, sería extraño publicar hipótesis como estas en un libro dirigido a un público general antes de someterlas a una revisión por pares y publicarlas en las revistas especializadas adecuadas. Lo mismo ocurre con las presentaciones orales, en las que debemos defender nuestros argumentos en foros científicos y académicos. He presentado las ideas de este libro a diversas audiencias de todo el mundo, en varias disciplinas especializadas.[469]

La suerte está echada. ¿Significa esto que debería haberme abstenido de publicar estos artículos y presentar estas ponencias? La respuesta es inequívocamente no. Si no lo hubiera hecho yo, lo habría hecho otro. Estas ideas están en el aire. Ha llegado su momento. No estoy a la defensiva; es una verdad demostrable. Tengamos en cuenta, por ejemplo, el artículo de nuestros colegas checos que predijeron en 2017 que una solución al problema difícil era inminente y que se basaría en el principio de la energía libre.[470] Robin Carhart-Harris (que asistió a algunas de las primeras reuniones neuropsicoanalíticas celebradas en Londres) ha publicado de forma independiente ideas que siguen líneas similares.[471] Lo mismo se puede decir de la neurocientífica social

Katerina Fotopoulou, que reconoció el vínculo entre la incertidumbre (precisión inversa) y la conciencia ya en 2013.[472] Sin lugar a dudas, si el problema difícil de la conciencia va a ser resuelto a través de alguna combinación de las cuatro ideas anteriormente enumeradas, es algo que sucederá en un futuro muy próximo, con o sin mi participación.

Esta constatación orienta el enfoque que he decidido dar a las cuestiones éticas y morales que estamos considerando. No me planteo retener publicaciones y otras presentaciones; en cambio, siempre que sea posible, quiero ayudar a la aplicación práctica de mis hipótesis, y hacerlo sin demora. Este planteamiento se deriva lógicamente de la expectativa de que si se puede hacer, se hará. Por lo tanto, debo intentar adelantarme a la ola, para estar en condiciones de prevenir las consecuencias potencialmente perjudiciales, siempre que sea posible.

Básicamente, esto significa que el proyecto esbozado en este capítulo debe ponerse en marcha ahora y debe llevarse a cabo sin ninguna financiación comercial. Suponiendo que mi equipo de investigación consiga alcanzar el criterio antes esbozado, a saber, la supervivencia del yo artificial en entornos impredecibles, y suponiendo que obtengamos pruebas razonables de que ese yo artificial es sintiente (es decir, si no se desmienten mis predicciones al respecto), entonces, en mi opinión, deberíamos proceder inmediatamente con los tres pasos siguientes.

En primer lugar, creo que debemos apagar la máquina y extraer su batería interna. Me doy cuenta de que esto aborda una de las preocupaciones éticas que he mencionado con anterioridad, pero creo que es lo correcto en primera instancia. Hay que recordar que una máquina como la que aquí se plantea no estará viva. No veo ninguna razón por la que apagar una máquina consciente que no está viva deba implicar su desaparición. Siempre debería ser posible volver a encenderla, y es de suponer que el agente consciente así revivido será idéntico al que fue apagado (utilizando la analogía biológica del sueño y la vigilia). Cabe señalar que este plan de apagar nuestra máquina es coherente con la cláusula de «Obligación de terminación» contenida en las Directrices Universales para la Inteligencia Artificial (2018), que es «la máxima declaración de responsabilidad de un sistema de IA».[473]

En segundo lugar, debemos iniciar el proceso de patentar el componente crítico de nuestro sistema de tercera fase —su correlato neuronal de la conciencia, sea lo que sea lo que resulte ser—, es decir,

aquello que nos permitió alcanzar el criterio que habíamos enunciado. No es posible patentar meras ecuaciones, así que no hay riesgo de que alguien lo haga mientras intentamos aplicarlas en la práctica. Sin embargo, dado que las ecuaciones ya son de dominio público, es imperativo que actuemos con rapidez para poder controlar su aplicación concreta antes de que lo haga nadie. Si llega el caso, es importante que la patente se registre a nombre de una organización sin ánimo de lucro adecuada —como Open AI o el Future of Life Institute—, y no de una persona o grupo de personas. Como mínimo, así se garantiza la toma de decisiones colectiva y aumentan las posibilidades de que las decisiones se tomen en interés del bien común.

En tercer y último lugar, si logramos cumplir nuestros criterios y se registra la patente, sus guardianes deberían organizar un simposio en el que se invite a científicos y filósofos de renombre y a otras partes interesadas a considerar las implicaciones y a hacer recomendaciones sobre el camino que seguir, incluyendo si la máquina sintiente debería volver a encenderse, cuándo y en qué condiciones, y quizá seguir desarrollándose. Es de esperar que esto lleve a la elaboración de una serie de directrices y limitaciones más amplias para el futuro desarrollo, explotación y multiplicación de la IA sintiente en general.

Creo que hay que decir estas cosas, por muy sorprendido que yo mismo esté de haber llegado a estas recomendaciones. Una vez cumplidos estos tres pasos, no debemos hacernos ilusiones sobre su fragilidad. El precedente de la energía nuclear y las armas atómicas está a la vista de todos. No queda más remedio que hacer lo que podamos, tan pronto como podamos, para reconocer la magnitud de las implicaciones derivadas de nuestra inminente capacidad para crear máquinas conscientes.

Ahora parece existir un criterio objetivo razonablemente claro de sintiencia. Es de esperar que esto altere nuestro comportamiento ético de forma más general, más allá de las estrechas cuestiones que surgen de la perspectiva de crear máquinas conscientes. En general, se considera que los sentimientos son condiciones necesarias y suficientes para despertar preocupaciones éticas. La comprensión científica de los sentimientos esbozada en este libro nos brinda así la oportunidad de reflexionar un poco más profundamente sobre el sufrimiento animal. He mencionado más de una vez cómo los avances en neurociencia afectiva de finales del siglo XX —es decir, la constatación de que lo necesario para los seres sintientes es poco más que un triángulo de

decisión en el mesencéfalo, algo que compartimos todos los vertebrados— alteraron la opinión de muchos científicos sobre lo que es y no es aceptable en la investigación con animales. Parece evidente que lo mismo debería aplicarse a la actitud del público hacia el bienestar de los animales en general. Por ejemplo, ¿cómo justificar la cría y el sacrificio a escala industrial de seres sintientes para comérselos? Al abordar esta cuestión, debemos tener en cuenta que la conciencia surge por grados, de modo que la supuesta sintiencia de una mosca o un pez no se puede equiparar a la de un ser humano. Sin embargo, no debemos olvidar que las ovejas, las vacas y los cerdos (tan presentes en los menús occidentales) son mamíferos como nosotros. Esto quiere decir que están sujetos a las mismas emociones básicas, como el MIEDO, el PÁNICO-DOLOR y el CUIDADO. Los mamíferos también tienen una corteza cerebral, lo que significa que son capaces —todos ellos, hasta cierto punto— de «recordar el futuro» conscientemente y abrirse camino a través de sus probabilidades.

A medida que avanza el siglo XXI, y a falta de un objetivo superior —si todo lo que somos es nuestra conciencia—, ¿qué otra cosa podemos hacer salvo intentar minimizar el sufrimiento? Ahora que tenemos una idea más clara de dónde puede haber sufrimiento, ¿qué más podríamos hacer con este conocimiento? La preservación y protección de la conciencia biológica no está ligada únicamente al destino de nuestra especie.

Teniendo en cuenta las cosas a las que pido renunciar en las páginas de este libro, con respecto a conceptos como la excepcionalidad humana, tal vez sea apropiado terminar con una breve reflexión sobre lo que podríamos ganar en nuestra autoconcepción a partir de estas desagradables revelaciones.

Sentir es una herencia valiosísima. Lleva inscrita la sabiduría de los tiempos: una herencia que se remonta al principio de la vida misma. Cuando la homeostasis acabó dando lugar a los sentimientos, lo más importante de esta nueva capacidad fue que nos permitió saber cómo estamos dentro de una escala biológica de valores. Las sensaciones y los sentimientos dan lugar a predicciones que se basan en las experiencias acumuladas en situaciones de importancia biológica de —literalmente— todos nuestros antepasados; nos permiten hacer lo que es mejor para nosotros, aunque no sepamos por qué lo hacemos. Ya hemos intentado imaginar qué pasaría si cada uno de nosotros tuviera que aprender de nuevo qué alimentos contienen grandes reservas de energía y si tuviéramos que descubrir por nosotros mismos lo que ocurre cuando saltamos desde un acantilado. Debido a los sentimientos y sensaciones no buscados que nos atraen hacia lo dulce

y nos hacen evitar las alturas, «sabemos» de forma precisa (en una primera aproximación) qué hacer y cuándo. Por ejemplo, sabemos qué hacer cuando los bebés lloran, cuando los depredadores atacan o cuando se interponen obstáculos frustrantes en nuestro camino. Este conocimiento innato (que se nos transmite explícitamente solo en forma de sentimientos) es lo que nos permite sobrevivir en los mundos altamente impredecibles en los que vivimos, donde los vehículos a motor circulan a toda velocidad a nuestro alrededor y el dióxido de carbono invade el aire.

Así pues, mientras abandonamos la ilusión familiar de que la conciencia fluye a través de nuestros sentidos y la idea errónea de que conciencia es sinónimo de comprensión, nos consolaremos con el hecho de que en realidad procede espontáneamente de nuestro interior más íntimo. Amanece en nuestro interior incluso antes de nacer. Desde los orígenes, nos guía una corriente constante de sentimientos, que fluyen de un manantial de intuición que no sabemos de dónde brota. Ninguno de nosotros individualmente conoce las causas, pero las sentimos. Los sentimientos son un legado que nos ha dejado toda la historia de la vida para prepararnos para las incertidumbres que se avecinan.

[431] Quizá los pulpos no estén de acuerdo. (El sistema de lóbulos verticales parece ser el equivalente en los cefalópodos de la corteza cerebral en los vertebrados).

[432] Este grupo empezó con Tristan Hromnik, Jonathan Shock y yo. Luego se fue ampliando con la incorporación de otros físicos, informáticos e ingenieros biomédicos (George Ellis, Rowan Hodson, Leen Remmelzwaal, Amit Mishra, Dean Rance, Dawie van den Heever y Julianne Blignaut), así como de los neuropsicólogos Joshua Martin, Aimee Dollman y Donne van der Westhuizen. El equipo sigue creciendo, aunque Hromnik ya no forma parte de él y Martin se ha trasladado a Berlín.

[433] Searle, 1980. Damasio (2018) adoptó una postura similar.

[434] Me refiero al argumento de los «qualia danzantes» de Chalmers, que se basa en el argumento de los «espectros invertidos» de Locke. Chalmers, 1995a,b, 2011.

[435] Chalmers, 1995a, pp. 214-215.

[436] Ibid., p. 215; la cursiva es mía.

[437] Ibid.; la cursiva es mía.

[438] He mantenido largas conversaciones con él sobre este tema y mi impresión es que, como mínimo, tiene una mentalidad abierta.

[439] Ethier et al., 2012; Hochberg et al., 2012; Collinger et al., 2013; Bouton et al., 2016; Capogrosso et al., 2016.

[440] Capogrosso et al., 2016, p. 284.

[441] Ibid.

[442] Abu-Hassan et al., 2019.

[443] Pasley et al., 2012.

[444] Nishimoto et al., 2011.

[445] Horikawa et al., 2013.

[446] Herff et al., 2015.

[447] Incluidos los procesos descendentes, como los que se abordan en el capítulo 10 en la parte que trata de «pensamiento» (como imaginar y soñar).

[448] Cf. Kurzweil, 2005.

[449] Solms y Turnbull, 2002, pp. 70-71; la segunda cursiva es mía.

[450] Véanse Solms, 1996, 1997b.

[451] El célebre físico Richard Feynman tenía la misma opinión sobre la comprensión mecanicista en general. No quiero decir que la ingeniería inversa de la conciencia resuelva por sí misma el problema. Es posible ensamblar algo, microprocesador a microprocesador, sin entenderlo. Lo que quiero decir es que si se entiende, debería poderse hacer ingeniería inversa con ello.

[452] Véase Reggia, 2013, para una revisión de investigaciones anteriores en este sentido.

[453] Dado que el enfoque del aprendizaje por refuerzo requiere un

criterio objetivo, puede ser este: la supervivencia del sistema en entornos impredecibles.

[454] Quiero aclarar que esa materialización puede simularse (al igual que todos los parámetros físicos descritos a continuación). Desde el punto de vista del sistema, no importa lo que ocurre realmente «fuera», solo lo que ocurre en el modelo en relación con la información que recibe del exterior. Por tanto, el equipo de investigación puede simular un entorno para que el sistema lo modele, y así lo haremos para las primeras generaciones de nuestro sistema propuesto. Proceder de otro modo llevaría mucho tiempo y sería francamente peligroso (consideremos, por ejemplo, el parámetro de sobrecalentamiento que se describe a continuación). Cuando digo que es más «realista» materializar físicamente el sistema, lo que quiero decir es que hay problemas de modelado que surgen con el movimiento físico (por ejemplo) que no surgen con el movimiento simulado, y esto podría resultar importante para un sistema realmente realista. Por esta y otras razones, las generaciones posteriores de nuestro sistema propuesto se encarnarán en robots.

[455] Es decir, codificaremos reflejos e instintos artificiales. También utilizaremos «algoritmos genéticos».

[456] Wikipedia (https://en.wikipedia.org/wiki/Artificial_consciousness) a 21 de marzo de 2020. [El texto corresponde a nuestra traducción de la versión inglesa reproducida por el autor (N. de la T.)]. Para una visión alternativa, véase Reggia, 2013: «El autor de esta revisión cree que ninguno de los estudios anteriores examinados, incluso cuando afirman lo contrario, ha proporcionado todavía un argumento convincente de cómo el enfoque estudiado conduciría finalmente a la instanciación de la conciencia artificial. Por otro lado, y más positivamente, todavía no se ha presentado ninguna prueba (incluidos los trabajos analizados en esta revisión) de que la conciencia instanciada de las máquinas no pueda ser posible algún día, una opinión compartida por otros autores».

[457] Solms y Turnbull, 2002, pp. 68-69.

[458] Véanse Colby, Watt y Gilbert, 1966; Weizenbaum, 1976. Un programa informático llamado «Eugene Goostman», que simula a un niño ucraniano de trece años, superó la prueba de Turing en un acto celebrado en 2014 en la Royal Society de Londres.

[459] Por ejemplo, como es bien sabido, los ordenadores pueden superar a los mejores jugadores humanos tanto de ajedrez como de go,

que es más difícil que el ajedrez.

[460] Haikonen, 2012.

[461] Resulta conmovedor, teniendo en cuenta los prejuicios que sufrió el propio Alan Turing (el diseñador de la prueba).

[462] Hace varios años, dije en una reunión celebrada en Viena entre psicoanalistas e ingenieros de IA que una forma de demostrar la conciencia artificial es buscar pruebas de psicopatología artificial: «En la medida en que los ingenieros consigan emular con precisión la mente humana, descubrirán que su modelo es propenso a ciertos tipos de disfunción. Casi dan ganas de utilizarlo como criterio de su éxito» (Solms, 2008).

[463] Mathur, Lau y Guo, 2011.

[464] Consideremos el caso extremo de las «empresas autónomas descentralizadas».

[465] Lin, Abney y Bekey, 2011.

[466] Bill Gates, Stephen Hawking y Elon Musk, por ejemplo, han expresado serias reservas sobre este tema.

[467] No pretendo que esta creencia por sí sola proporcione una justificación ética para hacerlo. El hecho de que alguien cometa un asesinato en algún lugar, algún día, no justifica que yo cometa un asesinato aquí y ahora. Sigan leyendo…

[468] Solms y Friston, 2018; Solms, 2019a; Solms, 2020b.

[469] He aquí una lista parcial: «Where does consciousness fit in the Bayesian brain?», 18º Congreso Internacional de Neuropsicoanálisis, University College de Londres, 2017; «How and why consciousness arises», Departamento de Física, Universidad de Ciudad del Cabo, 2017; «How and why consciousness arises», The Centre for Subjectivity Research, Universidad de Copenhague, 2017; «The self as feeling and memory», Universidad del Ruhr, Bochum, 2018; «The conscious id, the psychoanalytic process and the hard problem of consciousness», Departamento de Filosofía, Universidad de Nueva York, 2019; «Why and how consciousness arises», Departamento de Psiquiatría, Hospital Mount Sinai, Nueva York, 2019; «Why are we conscious? Lessons from neuroscience», Vermont College of Medicine, Burlington, 2019; «What is consciousness?», The Melbourne Brain Centre, Australia, 2019; «Consciousness itself», The Science of

Consciousness, Interlaken, Suiza, 2019; «Consciousness itself is affect», Escuela de Filosofía de Múnich, Burkardus Haus, Wurzburgo, Alemania, 2019; «The hard problem of consciousness», Departamento de Filosofía, Universidad de Ciudad del Cabo, 2019; «Consciousness is predictive work in progress», Hospital Ichilov, Tel Aviv, 2019; «Why and how consciousness arises», Italian Psychoanalytic Dialogues, Roma, 2020.

[470] Havlik, Kozakova y Horače, 2017.

[471] Carhart-Harris y Friston, 2010; Carhart-Harris et al., 2014; Carhart-Harris, 2018.

[472] Véase el capítulo 9, nota 16.

[473] www.linking-ai-principles.org/term/656. Véase The Public Voice Coalition, 2018.

Posfacio

Poco después de acabar el primer borrador de este libro, me invitaron a presentar su tesis principal en el Congreso anual de Ciencia de la Conciencia celebrado en Interlaken (Suiza, 2019), lo que me obligó a destilar la mayor parte de lo que han leído aquí en formato de conferencia plenaria. Podría ser útil poner fin a nuestro largo viaje resumiendo los trece puntos que empleé para redactar aquella conferencia.

(1) Johannes P. Müller, el gran fisiólogo del siglo XIX, pensaba que los organismos animados «contienen algún elemento no físico o se rigen por principios distintos a los de las cosas inanimadas». Sus alumnos (Helmholtz, Brücke, Du Bois-Reymond y Ludwig, entre otros) no estaban de acuerdo, porque estaban convencidos de que «las únicas fuerzas que están activas en el organismo son las fuerzas físicas y químicas comunes». A su vez, su discípulo Sigmund Freud intentó fundar sobre esa base una ciencia natural de la mente en la que la vida mental podía reducirse a «estados cuantitativamente determinados de partículas materiales especificables». Por falta de métodos, Freud no consiguió llevar a cabo su proyecto y lo abandonó en 1896.

(2) Un siglo más tarde (1994), el biólogo pionero Francis Crick declaró que «nosotros, nuestras alegrías y nuestras penas, nuestros recuerdos y nuestras ambiciones, nuestro sentido de la identidad personal y del libre albedrío solo son en realidad el comportamiento de un gran número de células nerviosas y las moléculas a ellas asociadas». Crick nos exhortó a volver a intentar descubrir los correlatos neuronales de la conciencia, intento que llevó a cabo él mismo, aunque, por desgracia, utilizó como su modelo la conciencia visual.

(3) Respondiendo a Crick, el filósofo David Chalmers sostuvo que la búsqueda de correlatos neuronales de la conciencia que había emprendido Crick era un problema «fácil» —más correlacional que causal— cuya solución podía explicar dónde, pero no por qué y cómo, surge la conciencia. Para Chalmers, el problema «difícil» de la conciencia era ¿cómo y por qué las actividades neurofisiológicas producen la experiencia de la conciencia? En su opinión (y también en la de Thomas Nagel, su predecesor filosófico), el problema giraba en torno al carácter de «algo que es como» propio de la experiencia: «Un organismo tiene estados de conciencia mental si y solo si hay algo que

es como ser ese organismo, algo que es como para el organismo». Por lo tanto, el problema difícil es este: ¿por qué y cómo la cualidad subjetiva de la experiencia surge de sucesos neurofisiológicos objetivos?

(4) Preguntar cómo las cosas objetivas producen cosas subjetivas es hablar sin fundamento, y se corre el riesgo de convertir el problema difícil en algo más difícil de lo que tiene que ser. La objetividad y la subjetividad son perspectivas observacionales, no causas y efectos. Los sucesos neurofisiológicos no pueden producir sucesos psicológicos de la misma forma que el relámpago no produce el trueno; son manifestaciones paralelas de un solo proceso subyacente. La causa subyacente del relámpago y el trueno es la electricidad, cuyos legítimos mecanismos explican a ambos. Del mismo modo, los fenómenos fisiológicos y psicológicos pueden reducirse a causas unitarias, pero no entre ellos.

(5) Solemos describir las causas subyacentes de los fenómenos biológicos en términos «funcionales», y, a su vez, los mecanismos funcionales pueden reducirse a leyes naturales. Por ejemplo, ¿cuál es el mecanismo de la visión? No obstante, Chalmers señala con razón que el mecanismo funcional de la visión no explica cómo es ver, y es así porque la visión no es una función intrínsecamente consciente. La ejecución de las funciones visuales (incluso las propias del ser humano, como leer) no tiene que sentirse. La percepción ocurre de buen grado sin conciencia de lo que se percibe, y el aprendizaje se da sin conciencia de lo que se aprende. En consecuencia, Chalmers acertaba al preguntar por qué la ejecución de esas funciones va acompañada de experiencia y por qué no tiene lugar todo el procesamiento de información «en la oscuridad», sin sensaciones interiores. Que la ciencia no pueda contestar a esta pregunta plantea la posibilidad de que la conciencia no forme parte de la matriz causal ordinaria del universo.

(6) Es razonable formular la pregunta de Chalmers para todas las funciones cognitivas, no solo para las visuales, pero no ocurre lo mismo con las funciones afectivas. ¿Cómo se puede tener un sentimiento sin sentirlo? ¿Cómo podemos explicar el mecanismo funcional del afecto sin explicar por qué y cómo nos lleva a experimentar algo? Incluso Freud coincidía en este punto: «La esencia de una emoción es sin duda que seamos conscientes de ella, esto es, que llegue a conocimiento de la conciencia. En consecuencia, la posibilidad del atributo de inconsciencia quedaría completamente excluida en lo que a las emociones, los sentimientos y los afectos respecta».

(7) Con estos antecedentes, resulta de máximo interés observar que el funcionamiento cortical va acompañado de conciencia solo si se lo «permite» el sistema reticular activador del tronco encefálico superior. Una lesión de apenas dos milímetros cúbicos en esta región suprime toda conciencia. Muchos piensan que eso ocurre porque el tronco del encéfalo modula el nivel cuantitativo de la conciencia, pero esa idea es insostenible. La conciencia generada por el tronco encefálico superior tiene un contenido cualitativo propio: el afecto. Dado que la conciencia cortical depende de la conciencia del tronco encefálico, el afecto resulta ser la forma fundacional de la conciencia. El sujeto sintiente está literalmente constituido por afecto.

(8) El afecto es una forma ampliada de la homeostasis, que es un mecanismo biológico básico que surgió naturalmente con la autoorganización. Los sistemas autoorganizados sobreviven porque ocupan estados limitados; no se dispersan. Este imperativo de supervivencia condujo poco a poco a la evolución de mecanismos dinámicos complejos que sustentan la intencionalidad. La mismidad de los sistemas autoorganizados les concede un punto de vista, y esto es crucial y explica por qué adquiere tanta importancia hablar de la subjetividad de tal sistema: las desviaciones respecto a sus estados viables los registra el sistema, para el sistema, como necesidades.

(9) El afecto sopesa hedónicamente las necesidades biológicas, de tal modo que las desviaciones crecientes y decrecientes de los puntos de estabilización homeostática (errores de predicción crecientes y decrecientes) se sienten como displacer y placer, respectivamente. Cada categoría de necesidad —de las que hay una gran variedad— tiene una cualidad afectiva propia, y cada una activa programas de acción previstos para devolver al organismo a sus límites viables. Estos estados activos —es decir, las respuestas intencionales a los estados afectivos— adoptan la forma de reflejos e instintos innatos, gradualmente complementados por el aprendizaje a través de la experiencia de acuerdo con la ley del afecto.[474] Que un organismo sienta fluctuaciones en sus propias necesidades permite elegir y, por lo tanto, apoya la supervivencia en contextos imprevistos. Esa es la función biológica de la experiencia.

(10) No es posible sentir todas las necesidades a la vez. Las prioriza el triángulo decisorio del mesencéfalo, donde las necesidades actuales (errores de predicción residuales cuantificados como energía libre) que convergen en la sustancia gris periacueductal son clasificadas en relación con las oportunidades actuales (desplegadas en forma de «mapa de saliencia» bidimensional en los tubérculos cuadrigéminos superiores). Esto activa programas de acción condicionados, que se

desarrollan en contextos previsibles en un esquema jerárquico profundo de predicciones (el modelo generativo del prosencéfalo expandido). Las acciones generadas por afectos priorizados son voluntarias; en consecuencia, más que a algoritmos preestablecidos, están sujetas a elecciones de aquí y ahora. Tales elecciones se sienten en la conciencia exteroceptiva, que contextualiza el afecto. Las elecciones se basan en la ponderación de la precisión fluctuante (también llamada «excitación», «modulación», «ganancia postsináptica») de las señales de error entrantes que las necesidades priorizadas convierten en salientes, mientras se reservan en una memoria de trabajo, con la finalidad de minimizar la incertidumbre (maximizar la confianza) en una predicción actual respecto al modo de satisfacer la necesidad. Eso es «reconsolidación». Como dijo Freud, «la conciencia surge en remplazo de la huella mnémica».

(11) Las elecciones fiables y acertadas resultan en ajustes a largo plazo de las predicciones sensoriomotoras. Así, la conciencia exteroceptiva es trabajo predictivo en curso, cuya finalidad es establecer predicciones cada vez más profundas (con mayor certeza, menos conscientes) respecto al modo en que se podrían resolver las necesidades. Esta consolidación a largo plazo —y la transición de los sistemas de memoria «declarativa» a «no declarativa»— requiere reducir la complejidad en el modelo predictivo para facilitar la generalizabilidad. Aspiramos al automatismo —confianza absoluta—, pero nunca lo alcanzamos por completo. En la medida en que fracasamos, padecemos sentimientos. Dado que nunca logramos predicciones libres de errores, la pulsión por defecto (cuando todo va bien) es la BÚSQUEDA, el compromiso proactivo con la incertidumbre con la finalidad de resolverla por anticipado. Cuando se prioriza ese afecto, se siente como curiosidad e interés por el mundo.

(12) Esos son los mecanismos causales de la conciencia —en sus dos manifestaciones, neurológica y psicológica—; qué aspecto tiene y cómo se siente. Las funciones subyacentes pueden reducirse a leyes naturales, como la ley de Friston.[475] La autoorganización se sustenta en estas leyes, que pueden explicar cómo y por qué combatir proactivamente la entropía (por ejemplo, el olvido) se siente como algo en la misma medida en que otras leyes científicas pueden explicar otras cosas naturales. La conciencia, siendo parte de la naturaleza, es matemáticamente tratable.

(13) Todos los sistemas de conciencia conocidos están vivos, pero no todos los sistemas vivos son conscientes. Asimismo, todos los sistemas vivos son autoevidenciables, pero no todos los sistemas autoevidenciables están vivos. Si el argumento que exponemos aquí es

correcto, en principio se puede construir un sistema autoevidenciable artificialmente consciente. La conciencia puede producirse. Así se harán realidad los sueños más locos de Helmholtz y de otros miembros de la Sociedad Física de Berlín. Sin embargo, debemos cuestionar nuestros motivos para hacerlo, aceptar la responsabilidad colectiva por las consecuencias potencialmente graves y proceder con extrema cautela.

[474] Ley del afecto: «Si un comportamiento va sistemáticamente seguido de recompensas, se incrementará, y si va sistemáticamente seguido de castigos, disminuirá».

[475] Ley de Friston: «Todas las cantidades que puedan cambiar (que sean parte del sistema) cambiarán para minimizar la energía libre».

APÉNDICE

EXCITACIÓN E INFORMACIÓN

En un libro autorizado sobre el tema de la excitación cerebral, Pfaff (2005, pp. 2-6) comenta lo siguiente:

Satisfaciendo la necesidad de una «fuente de energía» para el comportamiento, la excitación explica el inicio y la persistencia de un comportamiento motivado en una amplia variedad de especies. [...] La excitación, al alimentar mecanismos pulsionales, potencia el comportamiento, mientras que hay motivos e incentivos específicos que explican por qué un animal hace una cosa y no otra. [...] El Dictionary of Ethology no solo hace hincapié en la excitación en el contexto del ciclo sueño-vigilia; también remite al estado general de receptividad del animal tal como lo indica la intensidad de estimulación necesaria para desencadenar una reacción comportamental. La excitación «mueve al animal de un estado de inactividad a un estado en que está preparado para la acción». En el caso de la acción dirigida, Niko Tinbergen, uno de los fundadores de la etología, diría que la excitación proporciona la energía motora para un «patrón de acción fijo» en respuesta a un «estímulo de señal». El diccionario no deja de lado la neuropsicología, pues también aborda los niveles de excitación indicados por el electroencefalograma (EEG) cortical. [...] Generaciones enteras de científicos conductistas han teorizado y confirmado experimentalmente que un concepto como la excitación es necesario para explicar el inicio, la intensidad y la persistencia de las respuestas conductuales. La excitación ofrece la fuerza fundamental que hace activos y receptivos a animales y humanos para que ejecuten comportamientos instintivos o aprendidos hacia objetos meta. La resistencia de una respuesta aprendida depende de la excitación y la pulsión. Hebb vio un estado de activación generalizada como fundamental para la ejecución cognitiva óptima. Duffy va incluso más lejos invocando el concepto de «activación» para explicar una parte significativa de la conducta de un animal.

El análisis de Pfaff de los principales componentes sugiere que la proporción de conducta para una amplia gama de datos que pueden explicarse por la «excitación generalizada» está entre el 30 y el 45 por ciento.

[Duffy] anticipó que las medidas cuantitativas fisiológicas o físicas permitirían un enfoque matemático de este aspecto de la ciencia del comportamiento. [...] Cannon introdujo el sistema nervioso autónomo como un mecanismo necesario mediante el cual la excitación prepara al animal para la acción muscular. Teorías enteras de la emoción se basaron en la activación del comportamiento. [...] Malmo unificó todo

ese material citando evidencia electroencefalográfica y datos fisiológicos que acompañan a los resultados conductuales a la hora de establecer la activación y la excitación como componentes primarios que impulsan todos los mecanismos conductuales. […] Ese es el problema clásico de la excitación: ¿cómo las influencias internas y externas despiertan el cerebro y el comportamiento, ya sea en humanos o en otros animales, ya sea en el laboratorio o en entornos etológicos naturales? Es importante reformular y resolver este problema porque estamos tratando con la receptividad al entorno, uno de los requisitos elementales de la vida animal. También resulta especialmente oportuno reformular y resolver el problema ahora, porque hay nuevas herramientas neurobiológicas, genéticas e informáticas que permiten abordajes de los «estados conductuales» que hasta ahora no eran posibles. […] Explicar la excitación nos permitirá comprender los estados de comportamiento que subyacen a grandes cantidades de mecanismos de respuesta específicos. Realizar el análisis de muchos comportamientos a la vez no solo es estratégico: la elucidación de los mecanismos de los estados conductuales lleva a la comprensión del estado de ánimo y el temperamento. Dicho de otra manera, gran parte de la neurociencia del siglo XXI iba dirigida a explicar la particularidad de las conexiones específicas estímulo-respuesta. Ahora estamos en condiciones de revelar los mecanismos de clases enteras de respuestas bajo el epígrafe «control del estado». Lo más importante son los mecanismos que determinan el nivel de excitación. […] Cualquier definición verdaderamente universal de la excitación debe ser elemental y fundamental, primitiva e indiferenciada, y no ha de derivar de funciones superiores del sistema nervioso central (SNC). Tampoco puede estar limitada por condiciones o medidas particulares y temporales. Por ejemplo, no puede limitarse a explicar respuestas a solo una modalidad de estímulo. La actividad motora voluntaria y las respuestas emocionales también deberían incluirse. Por lo tanto, propongo la siguiente definición operacional, que es intuitivamente satisfactoria y que conducirá a mediciones cuantitativas precisas: «La excitación generalizada es mayor en un animal o ser humano que está: (S) más alerta a estímulos sensoriales de todo tipo, (M) más activo motrizmente y (E) más reactivo emocionalmente». Esta es una definición concreta de la fuerza más fundamental del sistema nervioso. […] Los tres componentes pueden medirse con precisión. […] Salta a la vista que existe una neuroanatomía de la excitación generalizada, neuronas cuyos patrones de disparos conducen a ella y genes cuya pérdida la alteran. Por ende, […] la excitación generalizada es el estado conductual producido por las vías de la excitación, sus mecanismos electrofisiológicos e influencias genéticas. El hecho de que esos mecanismos produzcan las

misma alerta sensorial (S), la misma reactividad motora (M) y la misma reactividad emocional (E) que se establece en nuestra definición afirma la existencia de una función de excitación generalizada y la exactitud de su definición operacional.

Pfaff prosigue: «Dado que la excitación del SNC depende de la sorpresa y es impredecible, su cuantificación adecuada depende de la matemática de la información» (p. 13; la cursiva es mía). Como explica Pfaff, la ecuación de Shannon (1948) convierte la información en algo mensurable:

Si un suceso dado es perfectamente regular, por ejemplo, el tictac de un metrónomo, el suceso siguiente (el próximo tic) no nos dice nada nuevo. Tiene una probabilidad de ocurrencia (p) sumamente alta en exactamente ese intervalo de tiempo. […] No tenemos ninguna incertidumbre sobre si, en un intervalo de tiempo dado, se producirá el tic. En la ecuación de Shannon, la información de un suceso dado está en proporción inversa a su probabilidad. Dicho de otra manera, a mayor incertidumbre sobre la ocurrencia de un suceso, más información se transmitirá, inherentemente, cuando el suceso ocurra. […] Cuando todos los sucesos de una serie son igualmente probables, la información alcanza su valor máximo. El desorden maximiza el flujo de información. Procedente de la termodinámica, el término técnico para el desorden en la ecuación de Shannon es entropía, y su símbolo para la entropía es H. […] El contenido de información inherente a algún suceso x es:

$$H(x) = p(x) \log 2 \ (1 / p(x))$$

donde p(x) es la probabilidad del suceso x.

Pfaff resume (pp. 19-20):

Para que un animal inferior o un ser humano se exciten, tiene que registrarse algún cambio en el entorno [interoceptivo o exteroceptivo]. Si hay un cambio, debe haber cierta incertidumbre acerca del estado del entorno. Desde un punto de vista cuantitativo, en la medida en que hay incertidumbre, la predictibilidad disminuye. Vistas estas consideraciones, podemos emplear [la ecuación de Shannon] para afirmar que cuanto menos predecible es el entorno y mayor es la entropía, más información hay disponible. La excitación del cerebro y del comportamiento, y los cálculos de información, están inseparablemente unidos.

En resumen, los estímulos desconocidos, inesperados, desordenados e inusuales (contenido alto de información) producen y sostienen respuestas de excitación (p. 23).

La teoría de la información ha estado acechando desde el principio tras las investigaciones de comportamiento y los datos neurofisiológicos. Primero, con una lógica clara y simple, consideremos qué hace falta para que un animal o un ser humano se despierte a la acción. Segundo, consideremos qué hace falta para reconocer un estímulo familiar (habituación) y prestar atención especial a estímulos nuevos. Tercero, desde el punto de vista de quien experimenta, la teoría de la información proporciona métodos para calcular el contenido significativo de los trenes de impulsos nerviosos y cuantificar la carga cognitiva de ciertas situaciones ambientales. Pueden formularse nuevas preguntas. ¿Cuánta distorsión de un campo de estímulos sensoriales se requiere para que haya novedad? ¿Qué clases de generalización de un tipo específico de estímulos están permitidas para un tipo dado de respuesta? El enfoque teórico de la información nos ayudará a convertir la combinación de la genética, la neurofisiología y el comportamiento en una ciencia cuantitativa. Podemos emplear las «matemáticas de la excitación» para ayudar a analizar los mecanismos neurobiológicos.

Pfaff concluye (pp. 138-145):

Los sistemas de excitación del SNC combaten heroicamente y de un modo muy especial la segunda ley de la termodinámica. Responden de manera selectiva a las situaciones ambientales que tienen una entropía inherentemente alta: un alto grado de incertidumbre y, en consecuencia, de contenido de información. Sin embargo, al responder, los sistemas de excitación del SNC reducen eficazmente la entropía comprimiendo toda esa información en una única respuesta legítima. […] La neurobiología de la excitación es la neurociencia del cambio, de la incertidumbre, de la impredecibilidad y de la sorpresa…, es decir, de la ciencia de la información. Hasta ahora, en todos los análisis de los mecanismos de excitación del SNC —neuroanatómicos, fisiológicos, genéticos y conductuales—, los conceptos de la teoría de la información han demostrado ser útiles. Las matemáticas de la información proporcionan maneras de clasificar las respuestas a los estímulos naturales. Las células nerviosas codifican probabilidades e incertidumbres, con el resultado de poder guiar el comportamiento en situaciones impredecibles. La excitación del SNC propiamente dicha depende totalmente del cambio, de la incertidumbre, de lo impredecible y de la sorpresa. Ese enorme

fenómeno llamado «habituación», una reducción de la amplitud de respuesta a la repetición del mismo estímulo, impregna la neurofisiología, la ciencia del comportamiento y la fisiología autónoma; y nos muestra cómo una reducción del contenido de información conduce a una reducción de la excitación del SNC. Así pues, podemos decir que la teoría de la excitación y la teoría de la información fueron hechas la una para la otra.

Es importante reconocer que las «matemáticas de la información» explican el comportamiento de las neuronas tanto en los procesos de excitación como de aprendizaje, los cuales, combinados, determinan lo que hace el cerebro. Por lo tanto, aunque la «información» no es un constructo psicológico, explica legítimamente la actividad fisiológica del cerebro. Esa es la función seleccionada por la evolución; los fenotipos fisiológicos vienen después.

AGRADECI-
MIENTOS

Quiero agradecer a los siguientes amigos y colegas la lectura de los sucesivos borradores de los capítulos del presente libro: Richard Astor, Nikolai Axmacher, Samantha Brooks, Aimee Dollman, George Ellis, Karl Friston, Eliza Kentridge (que es mucho más que una amiga), Joe Krikler, Joshua Martin, Lois Oppenheim, Jonathan Shock, Pippa Skotnes y Dawie van den Heever. Tengo una deuda especial con Ed Lake por dejar el manuscrito en una versión mucho más legible; nunca había visto trabajar tanto a un revisor. La edición final corrió a cargo de Trevor Horwood, con aportes colaterales de Tim James.

Asimismo quisiera dar las gracias a sir Sydney Kentridge por prestarme su casa de Chailey, donde escribí el grueso de este libro en los inviernos de 2018 a 2019 y de 2019 a 2020. Entre bastidores, como siempre, estuvieron mis intrépidas ayudantes Paula Barkay y Eleni Pantelis. Como ocurre con la mayor parte de todo lo que he hecho, sin ellas este libro hoy no estaría escrito.

BIBLIOGRAFÍA

Abbott, A., «What animals really think», Nature, 584, 2020, pp. 182-185.

Absher, J. y D. Benson, «Disconnection syndromes: an overview of Geschwind's contributions», Neurology, 43, 993, pp. 862-867.

Abu-Hassan, K., J. Taylor, P. Morris et al., «Optimal solid state neurons», Nature Communications, 10, 2019, p. 5309.

Adams, R., S. Shipp y K. Friston, «Predictions not commands: active inference in the motor system», Brain Structure and Function, 218, 2013, pp. 611-643.

Addis, D., A. Wong y D. Schacter, «Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration», Neuropsychologia, 45, 2017, pp. 1363-1377.

Ainley, V., M. A. J. Apps, A. Fotopoulou y M. Tsakiris, «"Bodily precision": a predictive coding account of individual differences in interoceptive accuracy», Philosophical Transactions of the Royal Society of London, B, 371, 2016 (1708), doi.org/10.1098/rstb.2016.0003.

Alboni, P., «Vasovagal syncope as a manifestation of an evolutionary selected trait», Journal of Atrial Fibrillation, 7, 2014, p. 1035.

Aserinsky, E. y N. Kleitman, «Regularly occurring periods of eye motility, and concomitant phenomena, during sleep», Science, 118, 1953, pp. 273-274.

Ashby, W., «Principles of the selforganizing dynamic system», Journal of General Psychology, 37, 1947, pp. 125-128.

Atkinson, R. y R. Shiffrin, «The control of short-term memory», Scientific American, 225, 1971, pp. 82-90.

Baars, B., A Cognitive Theory of Consciousness, Cambridge: Cambridge University Press, 1997.

—, In the Theatre of Consciousness, Oxford: Oxford University Press, 1997.

Bailey, P. y E. Davis, «The syndrome of obstinate progression in the cat», Experimental Biology and Medicine, 52, 1942, p. 307.

Bargh, J. y T. Chartrand, «The unbearable automaticity of being», American Psychologist, 54, 1999, pp. 462-479.

Barrett, L. F., How Emotions are Made: The Secret Life of the Brain, Nueva York: Houghton Mifflin Harcourt, 2017.

Bastos, A., W. Usrey, R. Adams et al., «Canonical microcircuits for predictive coding», Neuron, 76, 2012, pp. 695-711.

Bastos, A., J. Vezoli, C. Bosman et al., «Visual areas exert feedforward and feedback influences through distinct frequency channels», Neuron, 85, 2015, pp. 390-401.

Bayes, T., «An essay towards solving a problem in the doctrine of chances» [comunicado por el señor Price, en una carta a John Canton], Philosophical Transactions of the Royal Society of London, 53, 1763, pp. 370-418.

Bechtel, W. y R. Richardson, «Vitalism», en E. Craig (ed.), Routledge Encyclopedia of Philosophy, 9, Londres: Routledge, 1998, pp. 639-643.

Bentley, B., R. Branicky, C. Barnes et al., «The multilayer connectome of Caenorhabditis elegans», PLoS Computational Biology, 12, 2016, e1005283, doi.org/10.1371/journal.pcbi.1005283.

Berlin, H., «The neural basis of the dynamic unconscious», Neuropsychoanalysis, 13, 2011, pp. 5-31.

—, «The brainstem begs the question: "petitio principii"», Neuropsychoanalysis, 15, 2013, pp. 25-29.

Berridge, K., «Pleasures of the brain», Brain and Cognition, 52, 2003, pp. 106-128.

Besharati, S., S. J. Forkel, M. Kopelman, M. Solms, P. M. Jenkinson y A. Fotopoulou, «The affective modulation of motor awareness in anosognosia for hemiplegia: behavioural and lesion evidence», Cortex, 61, 2014, pp. 127-140.

Besharati, S., S. Forkel, M. Kopelman, M. Solms, P. Jenkinson y A. Fotopoulou, «Mentalizing the body: spatial and social cognition in anosognosia for hemiplegia», Brain, 139, 2016, pp. 971-985.

Besharati, S., A. Fotopoulou y M. Kopelman, «What is it like to be confabulating?», en A. L. Mishara, A. Kranjec, P. Corlett, P. Fletcher y

M. A. Schwartz (eds.), Phenomenological Neuropsychiatry, How Patient Experience Bridges Clinic with Clinical Neuroscience, Nueva York: Springer, 2014.

Bienenstock, E., L. Cooper y P. Munro, «Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex», Journal of Neuroscience, 2, 1982, pp. 32-48.

Blake, Y., D. Terburg, R. Balchin, J. van Honk y M. Solms, «The role of the basolateral amygdala in dreaming», Cortex, 113, 2019, pp. 169-183, doi.org/10.1016/j.cortex.2018.12.016.

Block, N., «On a confusion about a function of consciousness», Behavioral and Brain Sciences, 18, 1995, pp. 227-287.

Blomstedt, P., M. Hariz, A. Lees et al., «Acute severe depression induced by intraoperative stimulation of the substantia nigra: a case report», Parkinsonism and Related Disorders, 14, 2008, pp. 253-256.

Bogen, J., «On the neurophysiology of consciousness: 1. An overview», Consciousness and Cognition, 4, 1995, pp. 52-62.

Bouton, C., A. Shaikhouni, N. Annetta et al., «Restoring cortical control of functional movement in a human with quadriplegia», Nature, 533, 2016, pp. 247-250.

Bowlby, J., Attachment, Londres: Hogarth Press, 1969 [trad. cast.: El apego, Barcelona: Paidós, 1993, trad. de Mercedes Valcárcel].

Braun, A., «The new neuropsychology of sleep», Neuropsychoanalysis, 2, 1999, pp. 196-201.

Braun, A., T. Balkin, N. Wesenten et al., «Regional cerebral blood flow throughout the sleep-wake cycle. An H2(15)O PET study», Brain, 120, 1997, pp. 1173-1197.

Brentano, F., Psychologie vom empirischen Standpunkte, Leipzig: Duncker and Humblot, 1874 [trad. cast.: Psicología desde el punto de vista empírico, Salamanca: Sígueme, 2020, trad. de Sergio Sánchez-Migallón].

Broca, P., «Sur le principe des localisations cerebrales», Bulletin de la Société d’Anthropologie, 2, 1861, pp. 190-204.

—, «Sur le siège de la faculté du langage articulé», Bulletin de la Société d’Anthropologie, 6, 1865, pp. 377-393.

Brown, H., R. Adams, I. Parees, M. Edwards y K. Friston, «Active inference, sensory attenuation and illusions», Cognitive Processing, 14, 2013, pp. 411-427.

Cameron-Dow, C., «Do dreams protect sleep? Testing the Freudian hypothesis of the function of dreams», trabajo de fin de máster, University of Cape Town, 2012.

Campbell, A., «Histological studies on the localisation of cerebral function», Journal of Mental Science, 50, 1904, pp. 651-659.

Capogrosso, M., T. Milekovic, D. Borton et al., «A brain-spine interface alleviating gait deficits after spinal cord injury in primates», Nature, 539, 2016, pp. 284-288.

Carhart-Harris, R., «The entropic brain – revisited», Neuropharmacology, 142, 2018, pp. 167-178.

Carhart-Harris, R. y K. Friston, «The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas», Brain, 133, 2010, pp. 1265-1283.

Carhart-Harris, R., R. Leech, P. Hellyer et al., «The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs», Frontiers in Human Neuroscience, 8, 2014, artículo 20.

Chabris, C. y D. Simons, The Invisible Gorilla: and Other Ways Our Intuitions Deceive Us, Londres: Crown Publishers-Random House, 2010 [trad. cast.: El gorila invisible, Barcelona: RBA, 2011, trad. de Gabriela Ferrari].

Chalmers, D., «Facing up to the problem of consciousness», Journal of Consciousness Studies, 2, 1995 (a), pp. 200-219.

—, «Absent qualia, fading qualia, dancing qualia», en T. Metzinger (ed.), Conscious Experience, Paderborn: Ferdinand Schoningh, 1995 (b), pp. 309-328.

—, The Conscious Mind: In Search of a Fundamental Theory, Nueva York: Oxford University Press, 1996 [trad. cast.: La mente consciente. En busca de una teoría fundamental, Barcelona: Gedisa, 1999, trad. de José A. Álvarez].

—, «Consciousness and its place in nature», en S. Stich y T. Warfield (eds.), Blackwell Guide to the Philosophy of Mind, Londres: Blackwell,

2003, pp. 102-142.

—, «A computational foundation for the study of cognition», Journal of Cognitive Science, 12, 2011, pp. 325-359.

Charcot J.-M., «Un cas de suppression brusque et isolée de la vision mentale des signes et des objets (formes et couleurs)», Progres Medical, 11, 1883, p. 568.

Chew, Y., Y. Tanizawa, Y. Cho et al., «An afferent neuropeptide system transmits mechanosensory signals triggering sensitization and arousal in C. elegans», Neuron, 99, 2018, pp. 1233-1246.

Cisek, P. y J. Kalaska, «Neural mechanisms for interacting with a world full of action choices», Annual Review of Neuroscience, 33, 2010, pp. 269-298.

Claparède, É., «Recognition et moitié», Archives de Psychology, Genève, 11, 1911, pp. 79-90.

Clark, A., Surfing Uncertainty: Prediction, Action, and the Embodied Mind, Nueva York: Oxford University Press, 2015.

—, «Busting out: predictive brains, embodied minds, and the puzzle of the evidentiary veil», Nous, 51, 2017, pp. 727-753.

Coenen, A., «Consciousness without a cortex, but what kind of consciousness is this?», Behavioral and Brain Sciences, 30, 2007, pp. 87-88.

Coenen, V., B. Bewernick, S. Kayser et al., «Superolateral medial forebrain bundle deep brain stimulation in major depression: a gateway trial», Neuropsychopharmacology, 44, 1019, pp. 1224-1232, doi.org/10.1038/s41386-019-0369-9.

Colby, K., J. Watt y J. Gilbert, «A computer method of psychotherapy», Journal of Nervous and Mental Disease, 142, 1966, pp. 148-152.

Cole, S., A. Fotopoulou, M. Oddy y C. Moulin, «Implausible future events in a confabulating patient with an anterior communicating artery aneurysm», Neurocase, 20, 2014, pp. 208-224.

Collinger J., B. Wodlinger, J. Downey et al., «High-performance neuroprosthetic control by an individual with tetraplegia», The Lancet, 381, 2013, pp. 557-564.

Coltheart, M. y M. Turner, «Confabulation and delusion», en W. Hirstein (ed.), Confabulation: Views from Neuroscience, Psychiatry, Psychology and Philosophy, Nueva York: Oxford University Press, 2009, p. 173.

Conant, R. y W. Ashby, «Every good regulator of a system must be a model of that system», International Journal of Systems Science, 1, 1970, pp. 89-97.

Coren, S. y C. Porac, «The fading of stabilized images: Eye movements and information processing», Perception & Psychophysics, 16, 1974, pp. 529-534.

Corlett, P. y P. Fletcher, «Computational psychiatry: a Rosetta Stone linking the brain to mental illness», Lancet Psychiatry, 1, 2014, pp. 399-402.

Craig, A. D., «How do you feel – now? The anterior insula and human awareness», Nature Reviews Neuroscience, 10, 2019, pp. 59-70.

—, «Significance of the insula for the evolution of human awareness of feelings from the body», Annals of the New York Academy of Sciences, 1225, 2011, pp. 72-82.

Crick, F., The Astonishing Hypothesis: The Scientific Search for the Soul, Nueva York: Charles Scribner's Sons, 1994 [trad. cast.: La búsqueda científica del alma, Barcelona: Debate, 1994, trad. de Francisco Páez de la Cadena].

—, prólogo a C. Koch, The Quest for Consciousness: A Neurobiological Approach, Englewood, Colorado: Roberts and Company, 2004.

Crick, F. y C. Koch, «Towards a neurobiological theory of consciousness», Seminars in Neuroscience, 2, 1990, pp. 263-275.

Crucianelli, L., C. Krahe, P. Jenkinson y A. Fotopoulou, «Interoceptive ingredients of body ownership: affective touch and cardiac awareness in the rubber hand illusion», Cortex, 104, 2017, pp. 180-192, doi.org/10.1016/j.cortex.2017.04.018.

Cukur, T., S. Nishimoto, A. Huth y J. Gallant, «Attention during natural vision warps semantic representation across the human brain», Nature Neuroscience, 16, 2013, pp. 763-770.

Dahan, L., B. Astier, N. Vautrelle et al., «Prominent Burst Firing of Dopaminergic Neurons in the Ventral Tegmental Area during

Paradoxical Sleep», Neuropsychopharmacology, 32, 2007, pp. 1232-1241.

Damasio, A., Descartes’ Error: Emotion, Reason, and the Human Brain, Nueva York: Putnam, 1994 [trad. cast.: El error de Descartes, Barcelona: Crítica, 2010, trad. de Joandomènec Ros].

—, The Strange Order of Things: Life, Feeling, and the Making of Cultures, Londres: Penguin Random House, 2018 [trad. cast.: El extraño orden de las cosas, Barcelona: Destino, 2018, trad. de Joandomènec Ros].

Damasio, A. y G. Carvalho, «The nature of feelings: evolutionary and neurobiological origins», Nature Reviews Neuroscience, 14, 2013, pp. 143-152.

Damasio, A. y H. Damasio, Lesion Analysis in Neuropsychology, Nueva York: Oxford University Press, 1989.

Damasio, A., H. Damasio y D. Tranel, «Persistence of feelings and sentience after bilateral damage of the insula», Cerebral Cortex, 23, 2013, pp. 833-846.

Damasio, A., T. Grabowski, A. Bechara et al., «Subcortical and cortical brain activity during the feeling of self-generated emotions», Nature Neuroscience, 3, 2000, pp. 1049-1056.

Darwin, C., On the Origin of Species, Londres: John Murray, 1859 [trad. cast.: El origen de las especies, Madrid: Alianza Editorial, 2023, trad. de Dulcinea Otero].

—, The Expression of Emotions in Man and Animals, Londres: John Murray, 1872 [trad. cast.: La expresión de las emociones, Pamplona: Laetoli, 2009, trad. de Xavier Belles i Ros].

Davies, P., The Demon in the Machine: How Hidden Webs of Information are Solving the Mystery of Life, Londres: Allen Lane, 2019.

Debiec, J., V. Doyere, K. Nader y J. LeDoux, «Directly reactivated, but not indirectly reactivated, memories undergo reconsolidation in the amygdala», Proceedings of the National Academy of Sciences, 103, 2006, pp. 3428-3433.

Decety, J. y A. Fotopoulou, «Why empathy has a beneficial impact on others in medicine: unifying theories», Frontiers in Behavioral

Neuroscience, 8, 2015, p. 457.

Dehaene, S. y J.-P. Changeux, «Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness», PLoS Biology, 3, 2005, p. e141.

—, «Experimental and theoretical approaches to conscious processing», Neuron, 70, 2011, pp. 200-227.

Dehaene, S. y L. Naccache, «Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework», Cognition, 79, 2001, pp. 1-37.

Dement, W. y N. Kleitman, «The relation of eye movements during sleep to dream activity: an objective method for the study of dreaming», Journal of Experimental Psychology, 53, 1957, pp. 339-346.

Depaulis, A. y R. Bandler, The Midbrain Periaqueductal Gray Matter: Functional, Anatomical, and Neurochemical Organization, Nueva York: Plenum Press, 1991.

Ditchburn, R. y B. Ginsborg, «Vision with a stabilized retinal image», Nature, 170, 1952, pp. 36-37.

Domhoff, W., The Emergence of Dreaming: Mind-Wandering, Embodied Simulation, and the Default Network, Nueva York: Oxford University Press, 2017.

Du Bois-Reymond, E., Untersuchungen über thierische Elektricität, 2, Berlín: Reimer, 1848-1884.

Du Bois-Reymond, E. (ed.), Jugendbriefe von Emil Du Bois-Reymond an Eduard Hallmann, zu seinem hundertsten Geburtstag, dem 7. November 1918, Berlín: Reimer, 1918.

Dudai, Y., «The shaky trace», Nature, 406, 2000, pp. 686-687.

Edelman, G., The Remembered Present: A Biological Theory of Consciousness, Nueva York: Basic Books, 1990.

Edlow, B., E. Takahashi, O. Wu et al., «Neuroanatomic connectivity of the human ascending arousal system critical to consciousness and its disorders», Journal of Neuropathology and Experimental Neurology, 71, 2012, pp. 531-546.

Einstein, A., «Uber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt», Annalen der Physik, 17, 1905, pp. 132-148.

Eisenberger, N., «The neural bases of social pain: evidence for shared representations with physical pain», Psychosomatic Medicine, 74, 2012, pp. 126-135.

Ekman, P., W. Friesen, M. O'Sullivan et al., «Universals and cultural differences in the judgements of facial expressions of emotion», Journal of Personality and Social Psychology, 53, 1987, pp. 712-717.

Ellis, G. y M. Solms, Beyond Evolutionary Psychology: How and Why Neuropsychological Modules Arise, Cambridge: Cambridge University Press, 2018.

England, J., «Statistical physics of self-replication», Journal of Chemical Physics, 139, 2013, 121923, doi.org/10.1063/1.4818538.

Ethier, C., E. Oby, M. Bauman y L. Miller, «Restoration of grasp following paralysis through brain-controlled stimulation of muscles», Nature, 485, 2012, pp. 368-371.

Ezra, M., O. Faull, S. Jbabdi y K. Pattinson, «Connectivity based segmentation of the periaqueductal gray matter in human with brainstem optimized diffusion MRI», Human Brain Mapping, 36, 2015, pp. 3459-3471.

Feldman, H. y K. J. Friston, «Attention, uncertainty, and free-energy», Frontiers in Human Neuroscience, 4, 2010, pp. 215, doi.org/10.3389/fnhum.2010.00215.

Ferrarelli, F. y G. Tononi, «The thalamic reticular nucleus and schizophrenia», Schizophrenia Bulletin, 37, 2011, pp. 306-315.

Fischer, D., A. Boes, A. Demertzi et al., «A human brain network derived from coma-causing brainstem lesions», Neurology, 87, 2016, pp. 2427-2434.

Flechsig, P., «Developmental (mylogenetic) localisation of the cerebral cortex in the human subject», The Lancet, 2, 1901, pp. 1027-1029.

—, «Gehirnphsyiologie und Willenstheorien», Fifth International Psychology Congress, Roma, pp. 73-89, en G. von Bonin (ed.), Some Papers on the Cerebral Cortex, Springfield, Illinois: Charles C. Thomas, 1905, pp. 181-200.

Forrester, G., R. Davis, D. Mareschal et al., «The left cradling bias: an evolutionary facilitator of social cognition?», Cortex, 118, 2018, pp. 116-131, doi.org/10.1016/j.cortex.2018.05.011.

Fotopoulou, A., «False-selves in neuropsychological rehabilitation: the challenge of confabulation», Neuropsychological Rehabilitation, 18, 2008, pp. 541-565.

—, «Disentangling the motivational theories of confabulation», en W. Histein (ed.), Confabulation: Views from Neurology, Psychiatry, and Philosophy, Nueva York: Oxford University Press, 2009.

—, «The affective neuropsychology of confabulation and delusion», Cognitive Neuropsychiatry, 15, 2010 (a), pp. 38-63.

—, «The affective neuropsychology of confabulation and delusion», en R. Langdon y M. Turner (eds.), Confabulation and Delusion, Nueva York: Psychology Press, 2010 (b), pp. 38-63.

—, «Beyond the reward principle: consciousness as precision seeking», Neuropsychoanalysis, 15, 2013, pp. 33-38.

Fotopoulou, A. y M. Conway, «Confabulation pleasant and unpleasant», Neuropsychoanalysis, 6, 2004, pp. 26-33.

Fotopoulou, A., M. Conway, D. Birchall, P. Griffiths y S. Tyrer, «Confabulation: revising the motivational hypothesis», Neurocase, 13, 2007, pp. 6-15.

Fotopoulou, A., M. Conway y M. Solms, «Confabulation: motivated reality monitoring», Neuropsychologia, 45, 2007, pp. 2180-2190.

Fotopoulou, A., M. Conway, M. Solms, M. Kopelman y S. Tyrer, «Self-serving confabulation in prose recall», Neuropsychologia, 46, 2008 (a), pp. 1429-1441.

Fotopoulou, A., M. Conway, S. Tyrer, D. Birchall, P. Griffiths y M. Solms, «Is the content of confabulation positive? An experimental study», Cortex, 44, 2008 (b), pp. 764-772.

Fotopoulou, A., M. Solms y O. Turnbull, «Wishful reality distortions in confabulation: a case report», Neuropsychologia, 42, 2004, pp. 727-744.

Fotopoulou, A. y M. Tsakiris, «Mentalizing homeostasis: the social origins of interoceptive inference», Neuropsychoanalysis, 19, 2017,

pp. 3-76.

Frank, J., «Clinical survey and results of 200 cases of prefrontal leucotomy», Journal of Mental Sciences, 92, 1946, pp. 497-508.

—, «Some aspects of lobotomy (prefrontal leucotomy) under psychoanalytic scrutiny», Psychiatry, 13, 1950, pp. 45-52.

Frank, M., «Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism», Journal of Cognitive Neuroscience, 1, 2005, pp. 51-72.

Freud, S. (1883), «Einleitung in der Nervenpathologie», manuscrito inédito, Washington D. C.: Library of Congress, 1883.

— (1886), «Report on my studies in Paris and Berlin», Standard Edition of the Complete Psychological Works of Sigmund Freud, 1, Londres: Hogarth, pp. 1-15 [trad. cast.: «Informe sobre mis estudios en París y Berlín», Obras completas, Buenos Aires: Amorrortu Editores, 1976, trad. de José Luis Etcheverry. A partir de ahora solo se indicará entre corchetes el título en castellano del texto en la edición de Amorrortu. Si no se hace constar nada, es que esa obra no está incluida en dicha edición. (N. de la T.)].

— (1888), «Gehirn. I. Anatomie des Gehirns», en A. Villaret (ed.), Handwörterbuch der gesamten Medizin, 1. Stuttgart: Ferdinand Enke, pp. 684-691.

— (1891), On Aphasia, Nueva York: International Universities Press.

— (1893a), «Charcot», Standard Edition of the Complete Psychological Works of Sigmund Freud, 3, Londres: Hogarth, pp. 11-23.

— (1893b), «Some points for a comparative study of organic and hysterical motor paralyses», Standard Edition of the Complete Psychological Works of Sigmund Freud, 1, Londres: Hogarth, pp. 155-172 [«Algunas consideraciones con miras a un estudio comparativo de las parálisis motrices orgánicas e histéricas»].

— (1894), «The neuro-psychoses of defence», Standard Edition of the Complete Psychological Works of Sigmund Freud, 3, Londres: Hogarth, pp. 45-61 [«Las neuropsicosis de defensa»].

— (1895), «Studies on hysteria», Standard Edition of the Complete Psychological Works of Sigmund Freud, 2, Londres: Hogarth

[«Histeria»].

— (1900), «The interpretation of dreams», Standard Edition of the Complete Psychological Works of Sigmund Freud, 4 y 5, Londres: Hogarth [«La interpretación de los sueños»].

— (1901), «The psychopathology of everyday life», Standard Edition of the Complete Psychological Works of Sigmund Freud, 6, Londres: Hogarth [«Psicopatología de la vida cotidiana»].

— (1912), «A note on the unconscious in psychoanalysis», Standard Edition of the Complete Psychological Works of Sigmund Freud, 12, Londres: Hogarth, pp. 255-266 [«Nota sobre el concepto de lo inconsciente en psicoanálisis»].

— (1914), «On narcissism: an introduction», Standard Edition of the Complete Psychological Works of Sigmund Freud, 14, Londres: Hogarth, pp. 67-102 [«Introducción al narcisismo»].

— (1915a), «Instincts and their vicisitudes», Standard Edition of the Complete Psychological Works of Sigmund Freud, 14, Londres: Hogarth, pp. 117-140 [«Pulsiones y destinos de pulsión»].

— (1915b), «The unconscious», Standard Edition of the Complete Psychological Works of Sigmund Freud, 14, Londres: Hogarth, pp. 166-204 [«Lo inconsciente»].

— (1920), «Beyond the pleasure principle», Standard Edition of the Complete Psychological Works of Sigmund Freud, 18, Londres: Hogarth, pp. 7-64 [«Más allá del principio del placer»].

— (1923), «The ego and the id», Standard Edition of the Complete Psychological Works of Sigmund Freud, 19, Londres: Hogarth, pp. 12-59 [«El yo y el ello»].

— (1925), «A note upon "the mystic writing-pad"», Standard Edition of the Complete Psychological Works of Sigmund Freud, 16, Londres: Hogarth, pp. 227-232 [«Nota sobre la "pizarra mágica"»].

— (1940 [1939]), «An outline of psychoanalysis», Standard Edition of the Complete Psychological Works of Sigmund Freud, 23, Londres: Hogarth, pp. 144-207 [«Esquema del psicoanálisis»].

— (1950a [1895]), «Extracts from the Fliess papers», Standard Edition of the Complete Psychological Works of Sigmund Freud, 1, Londres: Hogarth, pp. 177-280 [«Fragmentos de la correspondencia con

Fliess»].

— (1950b [1895]), «Project for a scientific psychology», Standard Edition of the Complete Psychological Works of Sigmund Freud, 1, Londres: Hogarth, pp. 283-397 [«Proyecto de psicología»].

— (1994 [1929]), carta a Einstein, 1929, en I. Grubrich-Simitis (1995), «No greater, richer, more mysterious subject … than the life of the mind», International Journal of Psychoanalysis, 76, pp. 115-122.

Friston, K., «A theory of cortical responses», Philosophical Transactions of the Royal Society of London, B, 360, 2005, pp. 815-836.

—, «The Free Energy Principle: a rough guide to the brain?», Trends in Cognitive Sciences, 13, pp. 293-301.

—, «Life as we know it», Journal of the Royal Society Interface, 10, 2013, 20130475, doi.org/10.1098/rsif.2013.0475.

Friston, K., M. Breakspear y G. Deco, «Perception and selforganized instability», Frontiers in Computational Neuroscience, 6, 2012, p. 44.

Friston, K., F. Rigoli, D. Ognibene et al., «Active inference and epistemic value», Cognitive Neuroscience, 6, 2015, pp. 187-214.

Friston, K., P. Schwartenbeck, T. FitzGerald, M. Moutoussis, T. Behrens y R. Dolan, «The anatomy of choice: dopamine and decisionmaking», Philosophical Transactions of the Royal Society of London, B, 369, 2014, doi.org/10.1098/rstb.2013.0481.

Friston, K., T. Shiner, T. Fitzgerald, J. Galea, R. Adams, H. Brown, R. Dolan, R. Moran, K. Stephan y S. Bestmann, «Dopamine, affordance and active inference», PLoS Computational Biology, 8, 2012, e1002327.

Friston, K. y K. Stephan, «Free-energy and the brain», Synthese, 159, 2007, pp. 417-458.

Friston, K., K. Stephan, R. Montague y R. Dolan, «Computational psychiatry: the brain as a phantastic organ», Lancet Psychiatry, 1, 2014, pp. 148-158.

Frith, C., S. Blakemore y D. Wolpert, «Abnormalities in the awareness and control of action», Philosophical Transactions of the Royal Society of London, B, 355, 2000, pp. 1771-1788.

Galin, D., «Implications for psychiatry of left and right cerebral specialization: a neurophysiological context for unconscious processes», Archives of General Psychiatry, 31, 1974, pp. 572-583.

Garcia-Rill, E., «Bottom-up gamma and stages of waking», Medical Hypotheses, 104, 2017, pp. 58-62.

Gloor, P., «Role of the amygdala in temporal lobe epilepsy», en J. Aggleton (ed.), The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, Nueva York: Wiley-Liss, 1992, pp. 505-538.

Golaszewski, S., «Coma-causing brainstem lesions», Neurology, 87, 2016, p. 10.

Goodglass, H., «Norman Geschwind» (1926-1984), Cortex, 22, 1986, pp. 7-10.

Gosseries, O., C. Schnakers, D. Ledoux et al., «Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state», Functional Neurology, 26, 2011, pp. 25-30.

Gregory, R., «Perceptions as hypotheses», Philosophical Transactions of the Royal Society of London, B, 290, 1980, pp. 181-197.

Haikonen, P., Consciousness and Robot Sentience, Nueva Jersey: World Scientific, 2012.

Harding, D., On Having No Head, Londres: Sholland Trust, 1961 [trad. cast.: Vivir sin cabeza, Barcelona: Kairós, 1994, trad. de Swami Vit Nisheddha].

Harlow, J., «Passage of an iron rod through the head», Boston Medical and Surgical Journal, 39, 1868, pp. 389-393.

Hartmann. E., D. Russ, M. Oldfield, R. Falke y B. Skoff, «Dream content: effects of 1-DOPA», Sleep Research, 9, 1980, p. 153.

Hassin, R., J. Bargh, A. Engell y K. McCulloch, «Implicit working memory», Consciousness and Cognition, 18, 2009, pp. 665-678.

Havlik, M., E. Kozakova y J. Horače, «Why and how: the future of the central questions of consciousness», Frontiers in Psychology, 8, 2017, p. 1797, doi.org/10.3389/fpsyg.2017.01797.

Hebb, D., The Organization of Behavior: A Neuropsychological Theory, Nueva York: Wiley, 1949 [trad. cast.: La organización de la conducta, Madrid: Debate, 1985, trad. de Tomás del Amor].

Helmholtz, H. von, Handbuch der physiologischen Optik, 3, Leipzig: Voss, 1867.

—, «Goethes Vorahnungen kommender naturwissenschaftlicher Ideen», en Vortrage und Reden, 2, Brunswick: Friedrich Vieweg und Sohn, 1892, pp. 335-361.

Herff, C., D. Heger, A. de Pesters et al., «Brain-to-text: decoding spoken phrases from phone representations in the brain», Frontiers in Neuroscience, 9, 2015, p. 217, doi.org/10.3389/fnins.2015.00217.

Hering, E., «Der Raumsinn und die Bewegungen des Auges», en L. Hermann (ed.), Handbuch der Physiologie, 3, parte 1: Physiologie des Gesichtssinnes, Leipzig: Vogel, 1879, pp. 343-601.

Hesselmann, G., S. Sadaghiani, K. Friston y A. Kleinschmidt, «Predictive coding or evidence accumulation? False inference and neuronal fluctuations», PLoS One, 5(3), 2010, e9926, doi.org/10.1371/journal. pone.0009926.

Hobson, J. A., «REM sleep and dreaming: towards a theory of protoconsciousness», Nature Reviews Neuroscience, 10, 2009, pp. 803-813.

Hobson, J. A. y K. Friston, «Waking and dreaming consciousness: neurobiological and functional considerations», Progress in Neurobiology, 98, 2012, pp. 82-98.

—, «Consciousness, dreams, and inference: the Cartesian theatre revisited», Journal of Consciousness Studies, 21, 2014, pp. 6-32.

Hobson, J. A. y R. McCarley, «The brain as a dream state generator: an activation-synthesis hypothesis of the dream process», American Journal of Psychiatry, 134, 1977, pp. 1335-1348.

Hobson, J. A., R. McCarley y P. Wyzinski, «Sleep cycle oscillation: reciprocal discharge by two brainstem neuronal groups», Science, 189, 1975, pp. 55-58.

Hochberg L., D. Bacher, B. Jarosiewicz et al., «Reach and grasp by people with tetraplegia using a neurally controlled robotic arm», Nature, 485, 2012, pp. 372-375.

Hohwy, J., «Attention and conscious perception in the hypothesis testing brain», Frontiers in Psychology, 3, 2012, 96, doi.org/10.3389/fpsyg.2012.00096.

—, The Predictive Mind, Nueva York: Oxford University Press, 2013.

Holeckova, I., C. Fischer, M.-H. Giard et al., «Brain responses to subject's own name uttered by a familiar voice», Brain Research, 1082, 2006, pp. 142-152.

Holstege, G., J. Georgiadis, A. Paans et al., «Brain activation during human male ejaculation», Journal of Neuroscience, 23, 2003, pp. 9185-9193.

Horikawa, T., M. Tamaki, Y. Miyawaki y Y. Kamitani, «Neural decoding of visual imagery during sleep», Science, 340, 2013, pp. 639-642.

Hsieh, P-J. y P. Tse, «Illusory color mixing upon perceptual fading and filling-in does not result in "forbidden colors"», Vision Research, 46, 2006, pp. 2251-2258.

Hume, D., Philosophical Essays Concerning Human Understanding, Londres: A. Millar, 1748 [trad. cast.: Investigación sobre el conocimiento humano, Madrid: Alianza Editorial, 1980, trad. de Jaime de Salas].

Hurley, M., D. Dennett y R. Adams, Inside Jokes: Using Humor to Reverse-Engineer the Mind, Cambridge, Massachusetts: MIT Press, 2011.

Ingvar, D., «"Memory of the future": an essay on the temporal organization of conscious awareness», Human Neurobiology, 4, 1985, pp. 127-136.

Jackson, F., «Epiphenomenal qualia», Philosophical Quarterly, 32, 1982, pp. 127-136.

—, «Postscript on 'What Mary Didn't Know'», en P. Moser y J. Trout (eds.), Contemporary Materialism, Londres: Routledge, 1995, pp. 184-189.

Jaspers, K., General Psychopathology. Chicago: University of Chicago Press, 1963 [trad. cast.: Psicopatología general, Ciudad de México: Fondo de Cultura Económica, 2014, trad. de Roberto Saubidet].

Jaynes, E., «Information theory and statistical mechanics», Physical Review, 106, 1957, pp. 620-630.

Jouvet, M., «Paradoxical sleep: a study of its nature and mechanisms», Progress in Brain Research, 18, 1965, pp. 20-62.

Joyce, J., «Bayes' theorem», Stanford Encyclopedia of Philosophy, 2008.

Julesz, B., Foundations of Cyclopean Perception, Chicago: University of Chicago Press, 1971.

Kandel, E., «A new intellectual framework for psychiatry», American Journal of Psychiatry, 155, 1998, pp. 457-469.

—, «Biology and the future of psychoanalysis: a new intellectual framework for psychiatry revisited», American Journal of Psychiatry, 156, 1999 pp. 505-524.

Kant, I., «Kritik der Urteilskraft», Kants gesammelte Schriften, 5, Berlín: Walter de Gruyter, 1790.

Kaplan-Solms, K. y M. Solms, Clinical Studies in NeuroPsychoanalysis: Introduction to a Depth Neuropsychology, Londres: Karnac, 2000 [trad. cast.: Estudios clínicos en neuropsicoanálisis. Introducción a la neuropsicología profunda, Ciudad de México: Fondo de Cultura Económica, 2009].

Kihlstrom, J., «Perception without awareness of what is perceived, learning without awareness of what is learned», en M. Velmans (ed.), The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews, Londres: Routledge, 1996, pp. 23-46.

Knill, J. y A. Pouget, «The Bayesian brain: the role of uncertainty in neural coding and computation», Trends in Neurosciences, 27, 2004, pp. 712-719.

Koch, C., The Quest for Consciousness: A Neurobiological Approach, Englewood, Colorado: Roberts and Company, 2004.

Kopelman, M., A. Bajo y A. Fotopoulou, «Confabulation: memory deficits and neuroscientific aspects», en J. Wright (ed.), International Encyclopedia of Social and Behavioral Sciences, Nueva York: Elsevier, 2015.

Krahe, C., A. Springer, J. Weinman y A. Fotopoulou, «The social

modulation of pain: others as predictive signals of salience – a systematic review», Frontiers in Human Neuroscience, 7, 2013, p. 386.

Kurzweil, R., The Singularity is Near: When Humans Transcend Biology, Nueva York: Viking, 2005 [trad. cast.: La Singularidad está cerca, Berlín: Lola Books, 2021, trad. de Carlos García Hernández].

Lane, N., The Vital Question: Why is Life the Way It Is?, Londres: Profile, 2015 [trad. cast.: La cuestión vital. ¿Por qué la vida es como es?, Barcelona, Ariel, 2016, trad. de Joandomènec Ros].

Lavie, P., H. Pratt, B. Scharf, R. Peled y J. Brown, «Localized pontine lesion: nearly total absence of REM sleep», Neurology, 34, 1984, pp. 118-120.

LeDoux, J., The Emotional Brain, Nueva York: Simon and Schuster, 1996 [trad. cast.: El cerebro emocional, Barcelona: Planeta, 1999, trad. de Marisa Abdala].

—, «Psychoanalytic theory: clues from the brain», Neuropsychoanalysis, 1, 1999, pp. 44-49.

LeDoux, J. y R. Brown, «A higher-order theory of emotional consciousness», Proceedings of the National Academy of Science, 114, 2017, pp. e2016-e2025.

Lee, J., B. Everitt y K. Thomas, «Independent cellular processes for hippocampal memory consolidation and reconsolidation», Science, 304, 2004, pp. 839-843.

Lena, I., S. Parrot, O. Deschaux et al., «Variations in extracellular levels of dopamine, noradrenaline, glutamate, and aspartate across the sleep-wake cycle in the medial prefrontal cortex and nucleus accumbens of freely moving rats», Journal of Neuroscience Research, 81, 2005, pp. 891-899.

Leng, G., The Heart of the Brain: The Hypothalamus and Its Hormones, Cambridge, Massachusetts: MIT Press, 2018.

LeVay, S., The Sexual Brain, Cambridge, Massachusetts: MIT Press, 1993 [trad. cast.: El cerebro sexual, Madrid: Alianza Editorial, 1995, trad. de Eva Rodríguez Halffter].

Levine, J., «Materialism and qualia: the explanatory gap», Pacific Philosophical Quarterly, 64, 1983, pp. 354-361.

Libet, B., C. Gleason, E. Wright y D. Pearl, «Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): the unconscious initiation of a freely voluntary act», Brain, 106, 1983, pp. 623-642.

Lichtheim, L., «On aphasia», Brain, 7, pp. 433-484.

Liepmann, H., «Das Krankheitsbild der Apraxie (‘motorischen Asymbolie’) auf Grund eines Falles von einseitiger Apraxie», Monatsschrift fur Psychiatrie und Neurologie, 8, 1900, pp. 15-44.

Lightman, A., Searching for Stars on an Island in Maine, Nueva York: Pantheon, 2018.

Lin, P., K. Abney y G. Bekey, Robot Ethics, Cambridge, Massachusetts: MIT Press, 2011.

Linnman, C., E. Moulton, G. Barmettler et al., «Neuroimaging of the periaqueductal gray: state of the field», Neuroimage, 60, 2012, pp. 505-522.

Lisman, J. y G. Buzsaki, «A neural coding scheme formed by the combined function of gamma and theta oscillations», Schizophrenia Bulletin, 34, 2008, pp. 974-980.

Lissauer, H., «Ein Fall von Seelenblindheit, nebst einem Beitrag zur Theorie derselben», Archiv fur Psychiatrie und Nervenkrankheiten, 21, 1890, pp. 222-270.

Lupyan, G., «Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems», Review of Philosophy and Psychology, 6, 2015, pp. 547-569.

Lupyan, G. y A. Clark, «Words and the world: predictive coding and the language-perception-cognition interface», Current Directions in Psychological Science, 24, 2015, pp. 279-284.

Lupyan, G. y S. Thompson-Schill, «The evocative power of words: activation of concepts by verbal and nonverbal means», Journal of Experimental Psychology – General, 141, 2012, pp. 170-186.

Lupyan, G. y E. Ward, «Language can boost otherwise unseen objects into visual awareness», Proceedings of the National Academy of Sciences, 110, 2013, pp. 14196-14201.

Malcolm-Smith, S., S. Koopowitz, E. Pantelis y M. Solms, «Approach/

avoidance in dreams», Cognition and Consciousness, 21, 2012, pp. 408-412.

Man, K. y A. Damasio, «Homeostasis and soft robotics in the design of feeling machines», Nature Machine Intelligence, 1, 2019, pp. 446-452, doi.org/10.1038/s42256-019-0103-7.

Maturana, H. y F. Varela, Autopoiesis and Cognition: The Realization of the Living, Londres: Dordrecht, 1972.

Marc, J. y R. Llinás, «Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding», Proceedings of the National Academy of Sciences, 91, 1994, pp. 11748-11751.

Mathur, P., B. Lau y S. Guo, «Conditioned place preference behavior in zebrafish», Nature Protocols, 6, 2011, pp. 338-345.

Maxwell, J., Theory of Heat, Londres: Longmans, Green & Co, 1872.

Mazur, J. E., «Basic principles of operant conditioning», Learning and Behavior, 7.ª ed, Nueva York: Pearson, 2013, pp. 101-126.

McCarley, R. y J. A. Hobson, «The neurobiological origins of psychoanalytic dream theory», American Journal of Psychiatry, 134, 1977, pp. 1211-1221.

McKeever, W., «Tachistoscopic methods in neuropsychology», en H. J. Hannay (ed.), Experimental Techniques in Human Neuropsychology, Oxford: Oxford University Press, 1986, pp. 167-211.

Merker, B., «Consciousness without cerebral cortex: a challenge for neuroscience and medicine», Behavioral and Brain Sciences, 30, 2007, pp. 63-68.

Mesulam, M. M., «Behavioral neuroanatomy: large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations», en Principles of Behavioral and Cognitive Neurology, 2.ª ed, Nueva York: Oxford University Press, 2000, pp. 1-120.

Meynert, T., «Der Bau der Gross-Hirnrinde und seine ortliche Verschiedenheiten, nebst einem pathologisch-anatomischen Corollarium», Vierteljahrsschrift fur Psychiatrie in ihren Beziehungen zur Morphologie und Pathologie des Central-Nervensystems, der physiologischen Psychologie, Statistik und gerichtlichen Medizin, 1, 1867, pp. 77-93, 119-124.

—, Psychiatrie: Klinik der Erkrankungen des Vorderhirns, Viena: W. Braumüller, 1884.

Misanin, J., R. Miller y D. Lewis, «Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace», Science, 160, 1968, pp. 203-204.

Mohr, M. von y A. Fotopoulou, «The cutaneous borders of interoception: active and social inference of pain and pleasure on the skin», en M. Tsakiris y H. de Preester (eds.), The Interoceptive Basis of the Mind, Oxford: Oxford University Press, 2017.

Mongillo, G., O. Barak y M. Tsodyks, «Synaptic theory of working memory», Science, 319, 2008, pp. 1543-1546.

Montague, P., R. Dolan, K. Friston y P. Dayan, «Computational psychiatry», Trends in Cognitive Sciences, 16, 2012, pp. 72-80.

Moruzzi, G. y H. Magoun, «Brain stem reticular formation and activation of the EEG», Electroencephalography and Clinical Neurophysiology, 1, 1949, pp. 455-473.

Motta, S., A. Carobrez y N. Canteras, «The periaqueductal gray and primal emotional processing critical to influence complex defensive responses, fear learning and reward seeking», Neuroscience and Biobehavioral Reviews, 76, 2017, pp. 39-47.

Moustafa, A., S. Sherman y M. Frank, «A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism», Neuropsychologia, 46, 2008, pp. 3144-3156.

Mulert, C., E. Menzinger, G. Leicht et al., «Evidence for a close relationship between conscious effort and anterior cingulate cortex activity», International Journal of Psychophysiology, 56, 2005, pp. 65-80.

Munk, H., «Weiteres zur Physiologie des Sehsphare der Grosshirnrinde», Deutsche medizinische Wochenschrift, 4, 1878, pp. 533-536.

—, Über die Functionen der Grosshirnrinde: gesammelte Mittheilungen aus den Jahren 1877-80, Berlín: Albrecht Hirschwald, 1881.

Nader, K., G. Schafe y J. LeDoux, «Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval», Nature,

406, 2000, pp. 722-726.

Nagel, T., «What is it like to be a bat?», Philosophical Review, 83, 1974, pp. 435-450.

Newman, J. y B. Baars, «A neural attentional model for access to consciousness: a global workspace perspective», Concepts in Neuroscience, 4, 1993, pp. 255-290.

Niedenthal, P., «Embodying emotion», Science, 316, 2007, pp. 1002-1005.

Nishimoto, S., A. Vu, T. Naselaris et al., «Reconstructing visual experiences from brain activity evoked by natural movies», Current Biology, 21, 2011, pp. 1641-1646.

Nour, M. y R. Carhart-Harris, «Psychedelics and the science of self-experience», British Journal of Psychiatry, 210, 2017, pp. 177-179.

Nummenmaa, L., R. Hari, J. Hietanen y E. Glerean, «Maps of subjective feelings», Proceedings of the National Academy of Sciences, 115, 2018, pp. 9198-9203.

Oakley, D. y P. Halligan, «Chasing the rainbow: the non-conscious nature of being», Frontiers in Psychology, 8, 2017, p. 1924, doi.org/10.3389/ fpsyg.2017.01924.

Oberauer, K., A. Souza, M. Druey y M. Gade, «Analogous mechanisms of selection and updating in declarative and procedural working memory: experiments and a computational model», Cognitive Psychology, 66, 2013, pp. 157-211.

Okuda J., T. Fujii, H. Ohtake, T. Tsukiura, K. Tanji, K. Suzuki, R. Kawashima, H. Fukuda, M. Itoh y A. Yamadori, «Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes», Neuroimage, 19, 2003, pp. 1369-1380.

Pace-Schott, E. y J. A. Hobson, «Review of “The Neuropsychology of Dreams”», Trends in Cognitive Sciences, 2, 1998, pp. 199-200.

Paloyelis, Y., C. Krahe, S. Maltezos, S. Williams, M. Howard y A. Fotopoulou, «The analgesic effect of oxytocin in humans: a double-blind, placebo-controlled cross-over study using laser-evoked potentials», Journal of Neuroendocrinology, 28, 2016, 10.1111/ jne.12347.

Panksepp, J., «Hypothalamic regulation of energy balance and feeding behavior», Federation Proceedings, 33, 1974, pp. 1150-1165.

—, Affective Neuroscience: The Foundations of Human and Animal Emotions, Nueva York: Oxford University Press, 1998.

—, «The basic emotional circuits of mammalian brains: do animals have affective lives?», Neuroscience and Biobehavioral Reviews, 35, 2011, pp. 1791-1804.

Panksepp, J. y L. Biven, The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions, Nueva York: Norton, 2012.

Panksepp, J. y J. Burgdorf, «“Laughing” rats and the evolutionary antecedents of human joy», Physiology and Behavior, 79, 2003, pp. 533-547.

Panksepp, J. y M. Solms, «What is neuropsychoanalysis? Clinically relevant studies of the minded brain», Trends in Cognitive Sciences, 16, 2012, pp. 6-8.

Parr, T. y K. Friston, «The anatomy of inference: generative models and brain structure», Frontiers in Computational Neuroscience, 12, 2018, p. 90, doi.org/10.3389/fncom.2018.00090.

Partridge, M., PreFrontal Leucotomy: A Survey of 300 Cases Personally Followed for 1.-3 Years, Oxford: Blackwell, 1950.

Parvizi, J. y A. Damasio, «Neuroanatomical correlates of brainstem coma», Brain, 126, 2003, pp. 1524-1536.

Pasley, B., S. David, N. Mesgarani et al., «Reconstructing speech from human auditory cortex», PLoS Biology, 10(1), 2012, e1001251, doi.org/10.1371/journal.pbi0.1001251.

Pellis S. y V. Pellis, The Playful Brain: Venturing to the Limits of Neuroscience, Oxford: One World, 2009.

Penfield, W. y H. Jasper, Epilepsy and the Functional Anatomy of the Human Brain, Boston: Little, Brown & Company, 1954.

Petkova, V. y H. Ehrsson, «If I were you: perceptual illusion of body swapping», PLoS One, 3, 2008, e3832, doi.org/10.1371/journal.pone.0003832.

Pezzulo, G., «Why do you fear the bogeyman? An embodied predictive

coding model of perceptual inference», Cognitive Affective and Behavioral Neuroscience, 14, 2014, pp. 902-911.

Pfaff, D., Brain Arousal and Information Theory, Cambridge, Massachusets: Harvard University Press, 2005.

Picard, F. y K. Friston, «Predictions, perception, and a sense of self», Neurology, 83, 2014, pp. 1112-1118.

Popper, K., Conjectures and Refutations, Londres: Routledge, 1963 [trad. cast.: Conjeturas y refutaciones, Barcelona: Paidós, 1983, trad. de Rafael Grasa].

Qin, P., H. Di, Y. Liu et al., «Anterior cingulate activity and the self in disorders of consciousness», Human Brain Mapping, 31, 2010, pp. 1993-2002.

Ramachandran, V. S., «Filling in the blind spot», Nature, 356, 1992, p. 115.

Ramachandran, V. S. y R. Gregory, «Perceptual filling in of artificially induced scotomas in human vision», Nature, 350, pp. 699-702.

Ramachandran, V. S., R. Gregory y W. Aiken, «Perceptual fading of visual texture borders», Vision Research, 33, pp. 717-721.

Reggia, J., «The rise of machine consciousness: studying consciousness with computational models», Neural Networks, 44, 2013, pp. 112-131.

Riggs, L. y F. Ratliff, «Visual acuity and the normal tremor of the eyes», Science, 114, 1951, pp. 17-18.

Roepstorff, A. y C. Frith, «What's at the top in the top-down control of action? Script-sharing and "top-top" control of action in cognitive experiments», Psychological Research, 68, 2004, pp. 189-198.

Rolls, E., Emotion and DecisionMaking Explained, Nueva York: Oxford University Press, 2014.

—, «Emotion and reasoning in human decisionmaking», Economics, 13, 2019, pp. 1-31.

Rosenthal, D., Consciousness and Mind, Oxford: Oxford University Press, 2005.

Rovelli, C., Seven Brief Lessons in Physics, Londres: Allen Lane, 2014 [trad. cast.: Siete breves lecciones de física, Barcelona: Anagrama,

2017, trad. de F. J. Ramos Mena].

Runes, D., Dictionary of Philosophy, Totowa, Nueva Jersey: Littlefield, Adams and Co, 1972 [trad. cast.: Diccionario de filosofía, Barcelona: Grijalbo, 1985, trad. de Manuel Sacristán].

Sacks, O., Migraine, Londres: Vintage, 1972 [trad. cast.: Migraña, Barcelona: Anagrama, 2006, trad. de Gustavo Dessal y Damià Alou].

—, Awakenings, Londres: Duckworth, 1973 [trad. cast.: Despertares, Barcelona: Anagrama, 2010, trad. de Francesc Roca].

—, A Leg to Stand On, Nueva York: Simon and Schuster, 1984 [trad. cast.: Con una sola pierna, Barcelona: Anagrama, 2006, trad. de J. M. Álvarez Flórez].

—, The Man Who Mistook His Wife for a Hat, Londres: Duckworth, 1985 [trad. cast.: El hombre que confundió a su mujer con un sombrero, Barcelona: Anagrama, 2008, trad. de J. M. Álvarez Flórez].

Schacter, D., D. Addis y R. Buckner, «The prospective brain: remembering the past to imagine the future», Nature Reviews Neuroscience, 8, 2007, pp. 657-661.

Schindler, R., «Das Traumleben der Leukotomierten», Wiener Zeitschrift fur Nervenheilkunde, 6, 1953, p. 330.

Searle, J., «Minds, brains, and programs», Behavioral and Brain Sciences, 3, 1980, pp. 417-424.

—, The Rediscovery of the Mind, Cambridge, Massachusetts: MIT Press, 1992.

—, «The problem of consciousness», Social Research, 60, 1993, pp. 3-16.

—, The Mystery of Consciousness, Londres: Granta, 1997 [trad. cast.: El misterio de la conciencia, Barcelona: Paidós, 2000, trad. de Antoni Domènech].

Seeley, W., V. Menon, A. Schatzberg et al., «Dissociable intrinsic connectivity networks for salience processing and executive control», Journal of Neuroscience, 27, 2007, pp. 2349-2356.

Seth, A., «Interoceptive inference, emotion, and the embodied self», Trends in Cognitive Sciences, 17, 2013, pp. 565-573.

Shallice, T., From Neuropsychology to Mental Structure, Cambridge: Cambridge University Press, 1988.

Shannon, C., «A mathematical theory of communication», Bell System Technical Journal, 27, 1948, pp. 379-423.

Shapiro, L., Princess Elisabeth of Bohemia and René Descartes: The Correspondence Between Princess Elisabeth of Bohemia and René Descartes, Chicago: University of Chicago Press, 2007.

Sharf, B., C. Moskovitz, M. Lupton y H. Klawans, «Dream phenomena induced by chronic levodopa therapy», Journal of Neural Transmission, 43, 1978, pp. 143-151.

Shewmon, D., G. Holmes y P. Byrne, «Consciousness in congenitally decorticate children: developmental vegetative state as self-fulfilling prophecy», Developmental Medicine and Child Neurology, 41, 1999, pp. 364-374.

Skinner, B. F., Science and Human Behavior, Nueva York: Macmillan, 1953 [trad. cast.: Ciencia y conducta humana, Barcelona: Fontanella, 1987, trad. de M. J. Gallofré].

Solms, M., «Summary and discussion of the paper “The neuropsychological organisation of dreaming: implications for psychoanalysis”», Bulletin of the Anna Freud Centre, 16, 1991, pp. 149-165.

—, «New findings on the neurological organization of dreaming: implications for psychoanalysis», Psychoanalytic Quarterly, 64, 1995, pp. 43-67.

—, «Was sind Affekte?», Psyche, 50, 1996, pp. 485-522.

—, The Neuropsychology of Dreams: A Clinico-Anatomical Study, Mahwah, Nueva Jersey: Lawrence Erlbaum Associates, 1997 (a).

—, «What is consciousness?», Journal of the American Psychoanalytic Association, 45, 1997 (b), pp. 681-778.

—, «Before and after Freud’s “Project”», en R. Bilder y F. LeFever (eds.), «Neuroscience of the Mind on the Centennial of Freud’s Project for a Scientific Psychology», Annals of the New York Academy of Sciences, 843, 1998, pp. 1-10.

—, «Dreaming and REM sleep are controlled by different brain

mechanisms», Behavioral and Brain Sciences, 23, 2000 (a): pp. 843-850.

—, «Freud, Luria and the clinical method», Psychoanalysis and History, 2, 2000 (b), pp. 76-109.

—, «A psychoanalytic perspective on confabulation», Neuropsychoanalysis, 2, 2000 (c), pp. 133-138.

—, «The neurochemistry of dreaming: cholinergic and dopaminergic hypotheses», en E. Perry, H. Ashton y A. Young (eds.), The Neurochemistry of Consciousness, Nueva York: John Benjamins, 2001, pp. 123-131.

—, «What is the "mind"? A neuropsychoanalytical approach», en D. Dietrich, G. Fodor, G. Zucker y D. Bruckner (eds.), Simulating the Mind: A Technical Neuropsychoanalytical Approach, Viena: Springer Verlag, 2008, pp. 115-122.

—, «Neurobiology and the neurological basis of dreaming», en P. Montagna y S. Chokroverty (eds.), Handbook of Clinical Neurology, 98 (3.ª serie), Sleep Disorders, parte 1, Nueva York: Elsevier, 2011, pp. 519-544.

—, «The conscious id», Neuropsychoanalysis, 14, 2013, pp. 5-85 [trad. cast.: «El ello consciente», Revista Psicoanálisis, 20, Lima, 2017].

—, The Feeling Brain: Selected Papers on Neuropsychoanalysis, Londres: Karnac, 2015 (a).

—, «Reconsolidation: turning consciousness into memory», Behavioral and Brain Sciences, 38, 2015 (b), pp. 40-41.

—, «Empathy and other minds – a neuropsychoanalytic perspective and a clinical vignette», en V. Lux y S. Weigl (eds.), Empathy: Epistemic Problems and Cultural-Historical Perspectives of a Cross-Disciplinary Concept, Londres: Palgrave Macmillan, 2017 (a), pp. 93-114.

—, «Consciousness by surprise: a neuropsychoanalytic approach to the hard problem», en R. Poznanski, J. Tuszynski y T. Feinberg (eds.), Biophysics of Consciousness: A Foundational Approach, Nueva York: World Scientific, 2017 (b), pp. 129-148.

—, «What is "the unconscious", and where is it located in the brain? A neuropsychoanalytic perspective», Annals of the New York Academy

of Sciences, 1406, 2017 (c), pp. 90-97.

—, «Review of A. Damasio, "The Strange Order of Things"», Journal of the American Psychoanalytic Association, 66, 2018 (a), pp. 579-586.

—, «The scientific standing of psychoanalysis», British Journal of Psychiatry – International, 15, 2018 (b), pp. 5-8.

—, «The hard problem of consciousness and the Free Energy Principle», Frontiers in Psychology, 10, 2019 (a) p. 2714, doi.org/10.3389/fpsyg.2018.02714.

—, «Commentary on Edmund Rolls: "Emotion and reason in human decisionmaking"», Economics Discussion Papers, n.º 2019-45, Kiel: Institute for the World Economy, 2019 (b).

—, «Dreams and the hard problem of consciousness», en S. della Salla (ed.), Encyclopedia of Behavioral Neuroscience, Nueva York: Oxford University Press, pendiente de publicación.

—, «Notes on some technical terms whose translation calls for comment», en Solms, M. (ed.), Revised Standard Edition of the Complete Psychological Works of Sigmund Freud, 24, Lanham (Maryland): Rowman and Littlefield, pendiente de publicación.

—, «New project for a scientific psychology: general scheme», Neuropsychoanalysis, 21, 2020 (b).

Solms, M. y K. Friston, «How and why consciousness arises: some considerations from physics and physiology», Journal of Consciousness Studies, 25, 2018, pp. 202-238.

Solms, M., K. Kaplan-Solms y J. W. Brown, «Wilbrand's case of "mind-blindness"», en C. Code, C.-W. Walesch, A.-R. Lecours y Y. Joanette (eds.), Classic Cases in Neuropsychology, Hove: Erlbaum, 1996, pp. 89-110.

Solms, M., K. Kaplan-Solms, M. Saling y P. Miller, «Inverted vision after frontal lobe disease», Cortex, 24, 1988, pp. 499-509.

Solms, M. y J. Panksepp, «Why depression feels bad», en E. Perry, D. Collerton, F. LeBeau y H. Ashton (eds.), New Horizons in the Neuroscience of Consciousness, Ámsterdam: John Benjamins, 2010, pp. 169-179.

Solms, M. y M. Saling, «On psychoanalysis and neuroscience: Freud's

attitude to the localizationist tradition», International Journal of Psychoanalysis, 67, 1986, pp. 397-416.

—, A Moment of Transition: Two Neuroscientific Articles by Sigmund Freud, Londres: Karnac, 1990.

Solms, M. y O. Turnbull, The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience, Londres: Karnac, 2002 [trad. cast.: El cerebro y el mundo interior, Bogotá: Fondo de Cultura Económica, 2005].

—, «What is neuropsychoanalysis?» Neuropsychoanalysis, 13, 2011, pp. 133-145.

Solms, M. y M. Zellner, «Freudian drive theory today», en A. Fotopoulou, D. Pfaff y M. Conway (eds.), From the Couch to the Lab: Trends in Psychodynamic Neuroscience, Nueva York: Oxford University Press, 2012, pp. 49-63.

Squire, L., «The legacy of Patient HM for neuroscience», Neuron, 61, 2009, pp. 6-9.

Stein, T. y P. Sterzer, «Not just another face in the crowd: detecting emotional schematic faces during continuous flash suppression», Emotion, 12, 2012, pp. 988-996.

Stoerig, P. y E. Barth, «Low level phenomenal vision despite unilateral destruction of primary visual cortex», Consciousness and Cognition, 10, 2001, pp. 574-587.

Strawson, G., «Realistic monism – why physicalism entails panpsychism», Journal of Consciousness Studies, 13, 2006, pp. 3-31.

Sulloway, F., Freud: Biologist of the Mind, Nueva York: Burnett, 1979.

Szpunar, K., «Episodic future thought: an emerging concept», Perspectives on Psychological Science, 5, 2010, pp. 142-162.

Szpunar, K., J. Watson y K. McDermott, «Neural substrates of envisioning the future», Proceedings of the National Academy of Sciences, 104, 2007, pp. 642-647.

The Public Voice Coalition, Universal guidelines for Artificial Intelligence, borrador 9, 23 de octubre, Bruselas: Electronic Privacy Information Center, 2018, https://thepublicvoice.org/ai-universal-guidelines.

Thorndike, E., Animal Intelligence, Nueva York: Macmillan, 1911.

Tononi, G., «Integrated information theory of consciousness: an updated account», Archives of Italian Biology, 150, 2012, pp. 56-90.

Tossani, E., »The concept of mental pain», Psychotherapy and Psychosomatics, 82, 2013, pp. 67-73.

Tozzi, A., M. Zare y A. Benasich, «New perspectives on spontaneous brain activity: dynamic networks and energy matter», Frontiers in Human Neuroscience, 10, 2013, pp. 247, doi.org/10.3389/fnhum.2016.00247.

Tranel, D., G. Gullickson, M. Koch y R. Adolphs, «Altered experience of emotion following bilateral amygdala damage», Cognitive Neuropsychiatry, 11, 2006, pp. 219-232.

Turnbull, O., H. Berry y C. Evans, «A positive emotional bias in confabulatory false beliefs about place», Brain and Cognition, 55, 2004, pp. 490-494.

Turnbull, O., C. Bowman, S. Shanker y J. Davies, «Emotion-based learning: insights from the Iowa Gambling Task», Frontiers in Psychology, 5, 2014, pp. 162, doi.org/10.3389/fpsyg.2014.00162.

Turnbull, O., A. Fotopoulou y M. Solms, «Anosognosia as motivated unawareness: the “defence” hypothesis revisited», Cortex, 61, 2014, pp. 18-29.

Turnbull, O., S. Jenkins y M. Rowley, «The pleasantness of false beliefs: an emotion-based account of confabulation», Neuropsychoanalysis, 6, 2004, pp. 5-16.

Turnbull, O. y M. Solms, «Awareness, desire, and false beliefs», Cortex, 43, 2007, pp. 1083-1090.

Uhlhaas, P. y W. Singer, «Abnormal neural oscillations and synchrony in schizophrenia», Nature Reviews Neuroscience, 11, 2010, pp. 100-113.

Van der Westhuizen, D., J. Moore, M. Solms y J. van Honk, «Testosterone facilitates the sense of agency», Consciousness and Cognition, 56, 2017, pp. 58-67.

Van der Westhuizen, D. y M. Solms, «Social dominance in relation to the Affective Neuroscience Personality Scales», Consciousness and

Cognition, 33, 2015, pp. 90-111.

Varela, F., E. Thompson y E. Rosch, The Embodied Mind: Cognitive Science and Human Experience, Cambridge, Massachusetts: MIT Press, 1991.

Venkatraman, A., B. Edlow y M. Immordino-Yang, «The brainstem in emotion: a review», Frontiers in Neuroanatomy, 11, 2017, p. 15, doi.org/10.3389/fnana.2017.00015.

Vertes, R. y B. Kocsis, «Brainstem and diencephaloseptohippocampal systems controlling the theta rhythm of the hippocampus», Neuroscience, 81, 1997, pp. 893-926.

Walker, M., Why We Sleep, Londres: Penguin [trad. cast.: Por qué dormimos, Madrid: Capitán Swing Libros, 2019, trad. de Begoña Merino].

Wang, X. y J. Krystal, «Computational psychiatry», Neuron, 84, 2014, pp. 638-654.

Weiskrantz, L., Blindsight: A Case Study Spanning 35 Years and New Developments, Nueva York: Oxford University Press, 2009.

Weizenbaum, J., Computer Power and Human Reason: From Judgment to Calculation, Nueva York: W. H. Freeman, 1976.

Wernicke, C., Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis, Breslau: M. Crohn and Weigert, 1874.

Wheeler, A., A Journey into Gravity and Spacetime, Nueva York: W. H. Freeman, 1990 [trad. cast.: Un viaje por la gravedad y el espacio-tiempo, Madrid: Alianza Editorial, 1994, trad. de L. J. Garay].

White, B., D. Berg, J. Kan et al., «Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video», Nature Communications, 8, 2017, p. 14263, doi.org/10.1038/ncomms14263.

Whitty, C. y W. Lewin, «Vivid daydreaming; an unusual form of confusion following anterior cingulectomy», Brain, 80, 1957, pp. 72-76.

Wilbrand, H., Die Seelenblindheit als Herderscheinung und ihre Beziehungen zur homonymen Hemianopsie zur Alexie und Agraphie,

Wiesbaden: J. F. Bergmann, 1887.

—, «Ein Fall von Seelenblindheit und Hemianopsie mit Sectionsbefund», Deutsche Zeitschrift fur Nervenheilkunde, 2, 1892, p. 361.

Yang, E., D. Zald y R. Blake, «Fearful expressions gain preferential access to awareness during continuous flash suppression», Emotion, 7, 2007, pp. 882-886.

Yovell, Y., G. Bar, M. Mashiah et al., «Ultra-low-dose buprenorphine as a time-limited treatment for severe suicidal ideation: a randomized controlled trial», American Journal of Psychiatry, 173, 2016, pp. 491-498.

Yu, C. K.-C., «Cessation of dreaming and ventromesial frontal-region infarcts», Neuropsychoanalysis, 9, 2007, pp. 83-90.

Zahavi, D., «Brain, mind, world: predictive coding, neo-Kantianism, and transcendental idealism», Husserl Studies, 34, 2017, pp. 47-61, doi.org/10.1007/s10743-017-9218-z.

Zeki, S., A Vision of the Brain, Oxford: Blackwell, 1993 [trad. cast.: Una visión del cerebro, Barcelona: Ariel, 1995, trad. de Joan Soler Chic].

Zellner, M., D. Watt, M. Solms y J. Panksepp, «Affective neuroscientific and neuropsychoanalytic approaches to two intractable psychiatric problems: why depression feels so bad and what addicts really want», Neuroscience and Biobehavioral Reviews, 35, 2011, pp. 2000-2008.

Zeman, A., «Consciousness», Brain, 124, 2001, pp. 1263-1289.

Zhou, T., H. Zhu, Z. Fan et al., «History of winning remodels thalamo-PFC circuit to reinforce social dominance», Science, 357, 2017, pp. 162-168.

Índice

El manantial oculto

[image]

Para Mark Solms, uno de los pensadores más audaces de la neurociencia contemporánea, descubrir cómo surge la conciencia ha sido la búsqueda de toda una vida. Los científicos lo consideran el \"problema difícil\" porque parece una tarea imposible entender por qué sentimos un sentido subjetivo del yo y cómo surge en el cerebro.

Aventurándose en la física elemental de la vida, Solms ha llegado ahora a una respuesta asombrosa. En ‘El manantial oculto’, expone su descubrimiento en un lenguaje accesible y con analogías comprensibles.

Solms es un guía franco e intrépido en un viaje extraordinario desde los albores de la neuropsicología y el psicoanálisis hasta la vanguardia de la neurociencia contemporánea, ciñéndose a lo médicamente demostrable. Pero va más allá que otros neurocientíficos al prestar gran atención a las experiencias subjetivas de cientos de pacientes neurológicos, a muchos de los cuales trató, cuyas extrañas conversaciones ponen al descubierto muchas cosas sobre los oscuros alcances del cerebro.

Y lo que es más importante, usted será capaz de reconocer el funcionamiento de su propia mente por lo que realmente es, incluido cada pensamiento perdido, pulso de emoción y cambio de atención. ‘El Manantial Oculto’ alterará profundamente su comprensión de su propia experiencia subjetiva.

"Uno de los esfuerzos más meritorios surgidos de la neurociencia en los últimos tiempos". Jason Kehe, WIRED

"Nadie hechizado por estos misterios [de la conciencia] puede permitirse ignorar la solución propuesta por Mark Solms... Fascinante, amplio y sincero."- Oliver Burkeman, Guardian

"Sus ideas me parecen el futuro". Siri Hustvedt

"Remarcablemente claro, complaciente y emocionante de leer...’El manatial oculto’ proporciona un recordatorio necesario de que el pensamiento racional no es todo lo que parece ser". Jess Keiser, Washington Post

"Es un libro verdaderamente extraordinario. Lo cambia todo". Brian Eno

"Intrigante...Hay mucho para provocar y fascinar a lo largo del camino". Anil Seth, Times Higher Education

Mark Solms. Psicoanalista y neuropsicólogo sudafricano, conocido por su descubrimiento de los mecanismos cerebrales del sueño y su uso de métodos psicoanalíticos en la neurociencia contemporánea. Ocupa la Cátedra de Neuropsicología en la Universidad de Ciudad del Cabo y el Hospital Groote Schuur (Departamentos de Psicología y Neurología) y es Presidente de la Asociación Psicoanalítica Sudafricana. También es Presidente de Investigación de la Asociación Psicoanalítica Internacional (desde 2013). Fundó la Sociedad Internacional de Neuropsicoanálisis en 2000 y fue editor fundador (con Ed Nersessian) de la revista Neuropsychoanalysis. Es director del Centro Arnold Pfeffer de Neuropsicoanálisis del Instituto Psicoanalítico de Nueva York, director de la Fundación de Neuropsicoanálisis de Nueva York, fideicomisario del Fondo de Neuropsicoanálisis de Londres y director del Neuropsychoanalysis Trust de Ciudad del Cabo y fideicomisario del Loudoun Trust. Ha recibido numerosos premios, entre los que destacan el de Miembro Honorario de la Sociedad Psicoanalítica de Nueva York en 1998, del Colegio Americano de Psicoanalistas en 2004 y del Colegio Americano de Psiquiatras en 2015.

Título original: The Hidden Spring: A Journey to the Source of Consciousness (2022)

 Edición en ebook: mayo de 2024

c/ Rafael Finat 58, 2º 4 - 28044 Madrid Tlf: (+34) 630 022 531

28044 Madrid (España)

contacto@capitanswing.com

www.capitanswing.com

ISBN: 978-84-128388-4-8

Diseño de colección: Filo Estudio - www.filoestudio.com Corrección ortotipográfica: Victoria Parra Ortiz Composición digital: leerendigital.com